



**Politecnico
di Torino**

POLITECNICO DI TORINO

Collegio di Ingegneria Gestionale – Classe LM/31
Corso di Laurea Magistrale in Ingegneria Gestionale

Tesi di Laurea di II livello

PREVISIONE DATA-DRIVEN DELLO STATO DI USURA DEI VEICOLI DOTATI DI SISTEMI TELEMETRICI A SUPPORTO DEI PROCESSI DECISIONALI NELLA RIVENDITA DELL'USATO GARANTITO

Relatore:

Prof. Maurizio Galetto

Co-relatore:

Prof.ssa Elisa Verna

Candidato:

Nichole Natale

Anno Accademico 2025 - 2026

ABSTRACT

La presente tesi sviluppa un modello predittivo ibrido data-driven basato su K-Means clustering (k=5) e Isolation Forest, applicato a dati telemetrici OBD-II relativi al funzionamento motore di 3853 misurazioni mappati nel dataset obd2data, preso come riferimento del seguente studio.

L'applicazione degli algoritmi di Machine Learning su dati reali genera una classificazione oggettiva dello stato di usura (Ranking 1-5) dei veicoli monitorati, che sostituisce i paradigmi soggettivi di analisi tradizionalmente utilizzati nel settore della rivendita di auto usate.

L'elemento di innovazione del presente studio consiste nell'eliminazione dell'asimmetria informativa, intrinseca nel mercato della rivendita di auto usate, fornendo un metodo di valutazione oggettivo del prodotto, con un impatto atteso sul piano economico e ambientale.

La classificazione oggettiva del livello di usura delle auto rivendute sul mercato dell'usato si traduce dal punto di vista economico in un aumento del margine lordo per i rivenditori, il quale varia in base al ranking dell'auto venduta, ma che si stima in media pari a +15,6%. L'impatto economico sul mercato italiano della rivendita di auto non è dato soltanto dai superiori margini di guadagno, ma l'utilizzo di un metodo quantitativo fornisce ai consumatori maggiore certezza sul prodotto acquistato, aumentando il market share e generando nel mercato dell'usato garantito un aumento annuo complessivo di circa €2,5 miliardi.

L'applicazione del modello predittivo proposto aumenta l'affidabilità del prodotto modificando le abitudini di acquisto del cliente. In tal senso l'innovazione proposta genera un impatto ambientale in quanto attraverso il calcolo dell'usura dell'auto contribuisce a stimarne la vita utile residua e supporta il cliente nel preferire scelte di acquisto sostenibile, disincentivando l'acquisto di una nuova auto, stimando una riduzione di emissioni di CO₂ in Italia di circa 1,75-2,34 milioni di tonnellate annue, allineandosi con gli European Green Deal 2050.

Parole chiave: telemetrica OBD-II, K-Means unsupervised, economia circolare mobilità, asimmetria informativa automotive, sostenibilità ESG data-driven.

INDICE

CAPITOLO 1 - INTRODUZIONE E OBIETTIVI

1.1 Contesto generale	1
1.2 Motivazioni della ricerca	3
1.3 Obiettivo della tesi	4

CAPITOLO 2 - LA TELEMETRIA

2.1 Definizione e principi	6
2.2 Campi di applicazione	7
2.2.1 Monitoraggio remoto dei sistemi complessi	7
2.2.2 Ambito medico	7
2.2.3 Settore automotive	8
2.2.4 Risparmio energetico	8
2.2.5 Gestione e studio della fauna	8
2.2.6 Monitoraggio di macchinari industriali da remoto	9
2.3 Telemetria automobilistica	9
2.4 Trasformazione dell'esperienza di guida	11
2.5 Normative sulla sicurezza	13

CAPITOLO 3 – SENSORISTICA DI BORDO

3.1 Diagnostica del veicolo	15
3.2 Architettura telemetria automobilistica	16
3.3 Sottosistemi e parametri monitorati	17
3.3.1 Motore e sistema di alimentazione	18
3.3.2 Sistema di trasmissione e frizione	20

3.3.3 Sistema di frenata	21
3.3.4 Pneumatici e dinamica del veicolo	21
3.3.5 Sistema elettrico ed elettronico	22
3.3.6 Sospensioni e telaio	23
3.3.7 Sistema di scarico e controllo emissioni	24

CAPITOLO 4 – ANALISI DEL DATASET

4.1 Selezione del dataset	27
4.2 Descrizione del dataset	28
4.3 Studio del dataset	31
4.3.1 Natura del dataset	32
4.3.1.1 Distribuzione regime motore	33
4.3.1.2 Relazione tra regime motore e carico motore	35
4.3.1.3 Relazione tra principali parametri del dataset	36
4.3.2 Bilanciamento del dataset	39
4.3.3 Statistiche del dataset	40
4.3.3.1 Statistiche descrittive dei parametri	41
4.3.3.2 Test di normalità	43
4.3.3.3 Analisi di autocorrelazione	44
4.3.3.4 Analisi scree plot	45
4.3.3.5 Analisi outlier	46
4.3.3.6 Analisi della varianza	48

CAPITOLO 5 – ALGORITMO MACHINE LEARNING

5.1 Ribilanciamento del dataset.	50
5.2 Selezione del modello predittivo.	51

5.3 Scelta della scala	52
5.3.1 Elbow method	53
5.3.2 Silhouette score	55
5.4 Applicazione algoritmo	56
5.5 Analisi dei risultati	58

CAPITOLO 6 – APPLICAZIONI DELLO STUDIO

6.1 Implicazioni sulle decisioni di rivendita	61
6.2 Determinazione del prezzo di rivendita	62
6.3 Garanzia dell'usato	64
6.4 Impatto economico sul mercato dell'usato garantito	65
6.4.1 Aumento del margine dei rivenditori	65
6.4.2 Aumento del market share	66
6.5 Sostenibilità	67

CONCLUSIONI

BIBLIOGRAFIA

ELENCO FIGURE

Figura 1 – Istogramma distribuzione parametro regime motore.....	34
Figura 2 – Grafico di dispersione tra regime motore e carico motore	35
Figura 3 – Matrice di correlazione tra parametri del dataset	38
Figura 4 – Bilanciamento dataset.....	40
Figura 5 – Statistiche descrittive parametri dataset	41
Figura 6 – Test di normalità	43
Figura 7 – Analisi di autocorrelazione	45
Figura 8 – Scree Plot	46
Figura 9 – Analisi outliers	47
Figura 10 – Analisi varianza	48
Figura 11 – Metodo elbow	54
Figura 12 – Silhouette score	55
Figura 13 – Classificazione usura veicoli	57

ELENCO TABELLE

Tabella 1 – Principali driver di crescita della telemetria automobilistica	10
Tabella 2 – Principali protocolli presenti nelle reti automobilistiche	16
Tabella 3 – Parametri motore	19
Tabella 4 – Parametri trasmissione e frizione	20
Tabella 5 – Parametri frenata	21
Tabella 6 – Parametri pneumatici e dinamica del veicolo	22
Tabella 7 – Parametri elettrici ed elettronici	23
Tabella 8 – Parametri sospensioni e telaio	24
Tabella 9 – Parametri scarico ed emissioni	25
Tabella 10 – Parametri dataset obd2data	29
Tabella 11 – Ranking usura veicolo	53
Tabella 12 – Stima del prezzo di rivendita.	62
Tabella 13 – Stima dei livelli di garanzia	64

CAPITOLO 1

INTRODUZIONE E OBIETTIVI

1.1 Contesto generale

Negli ultimi anni il settore automotive è stato al centro di una rapida e profonda trasformazione, guidata dallo sviluppo dei sistemi di telemetria e dalla crescente adozione di sensori di bordo e tecnologie digitali avanzate all'interno dei veicoli. La loro integrazione all'interno della vettura ha portato un cambiamento significativo non solo nell'ambito della produzione automobilistica, ma anche nel modo in cui l'auto viene percepita e vissuta dal consumatore. Al giorno d'oggi, infatti, l'auto non è più soltanto un mezzo di trasporto, ma un'autentica esperienza di guida caratterizzata da aspetti che migliorano notevolmente il comfort dei passeggeri, la sicurezza nella guida e la gestione degli aspetti tecnici del veicolo. In tal senso, l'integrazione dei sistemi telemetrici nella vettura consente di raccogliere una grande quantità di dati relativi al funzionamento dei sottosistemi meccanici, elettronici e dinamici del veicolo, consentendo di monitorare in modo continuo lo stato di salute, le prestazioni e il comportamento dell'auto in condizioni operative reali [3][4].

In questo contesto di profonda innovazione che sta caratterizzando il settore automotive è emersa una nuova tendenza particolarmente interessante per le case automobilistiche, che pone al centro del quadro strategico la rivendita delle auto come usato garantito. Nel 2024, infatti, il mercato mondiale delle auto usate è stato stimato attorno ai 2.5 trilioni di Dollari Statunitensi (USD) e si prevede che lo stesso raggiungerà i 4 trilioni di Dollari Statunitensi (USD) nel 2034, con un Tasso di Crescita Annuale Composto (CAGR) del 4,5% [1].

I veicoli pre-owned si attestano come alternativa sempre più apprezzata dai consumatori budget-conscious, in quanto il rapido deprezzamento delle vetture di prima immatricolazione consente l'acquisto di modelli relativamente recenti ad un prezzo sensibilmente inferiore rispetto al loro costo originale. Alla luce delle nuove preferenze di acquisto dei consumatori, le case automobilistiche hanno introdotto programmi dedicati all'usato certificato che prevedono la possibilità di acquistare veicoli di seconda mano conformi agli standard di qualità definiti dal costruttore.

Tali vetture, infatti, prima di essere reimmesse sul mercato, vengono sottoposte ad un check-up approfondito che prevede oltre 150 controlli volti a testare i principali sottosistemi del veicolo, andando in questo modo a mitigare il rischio di informazione asimmetrica tra venditore e acquirente [3][5][7][8]. Una valutazione accurata dello stato della vettura e del suo livello di usura è essenziale per determinare il valore residuo dell'auto, pianificare eventuali interventi di ricondizionamento e ridurre l'incertezza economica legata alla rivendita. Tuttavia, le attuali procedure di controllo si basano principalmente su verifiche manuali e valutazioni peritali, la quale affidabilità risulta essere limitata dal fatto che si tratta di osservazioni puntuali, il cui esito dipende fortemente dal momento in cui è eseguita l'ispezione [3][7].

In questo contesto, l'integrazione tra sistemi telemetrici presenti nelle auto e tecniche avanzate di analisi dei dati apre prospettive promettenti per lo sviluppo di soluzioni che siano in grado di stimare in modo oggettivo il livello di usura dell'auto a partire dai dati reali raccolti dai sensori che monitorano il comportamento dei sottocomponenti dell'auto [3][4]. Sviluppare un approccio data-driven permetterebbe di valorizzare un patrimonio informativo altrimenti sottoutilizzato per ottenere una stima più robusta ed accurata dello stato di salute della vettura, migliorando l'affidabilità della diagnosi. Tale approccio, inoltre, ridurrebbe il numero di controlli manuali necessari prima della rivendita, sostituendoli con l'analisi dei dati raccolti dai sensori durante tutto il periodo di utilizzo dell'auto.

La telemetria automobilistica diventa quindi un elemento chiave non solo per il monitoraggio del veicolo, ma anche per la creazione di sistemi decisionali avanzati in grado di supportare il processo di valutazione dell'usato. L'analisi di dati reali applicata al settore della rivendita di auto usate contribuirebbe a diminuire l'asimmetria informativa tra acquirente e compratore, fornendo uno strumento oggettivo per determinare il valore residuo della vettura [3][4][5].

L'obiettivo centrale di questo lavoro di tesi è dimostrare come il vasto patrimonio informativo raccolto dai sistemi di telemetria di bordo possa essere reso funzionale allo sviluppo di metodologie per la diagnostica predittiva dello stato di usura dell'auto, supportando i processi di rivendita della stessa come usato garantito.

1.2 Motivazioni della ricerca

Di fronte alla crescente domanda di veicoli usati emerge l'esigenza di un approccio innovativo nella valutazione del livello di usura dell'auto che consenta una classificazione oggettiva delle vetture rivendute come usato garantito [1][7]. Le modalità di valutazione attualmente diffuse nel settore sono ancora fortemente ancorate ad ispezioni manuali che risultano essere dispendiose in termini di tempo, fortemente soggette ad errori e generalmente inadeguate per un'analisi completa [7][8].

Il processo di valutazione dell'usura reale di un veicolo rappresenta un problema complesso che coinvolge una molteplicità di fattori tecnici, comportamentali e operativi, in quanto risultato dell'interazione tra condizioni di utilizzo, sollecitazioni meccaniche ripetute, cicli di carico variabili e comportamenti di guida del conducente. A causa di questa complessità, le procedure di controllo attualmente impiegate nei programmi di usato certificato non permettono di avere una valutazione accurata dello stato di salute complessivo del veicolo. Trattandosi infatti di controlli puntuali, questi si limitano a restituire una fotografia dello stato del veicolo in quel momento, ma non permettono di risalire alle cause profonde delle anomalie rilevate o di ricostruire in modo completo il processo di degradazione a cui il veicolo è stato sottoposto nel tempo [7].

In questo contesto, i sistemi telemetrici di bordo rappresentano una fonte informativa estremamente ricca e ancora in larga parte sottoutilizzata. Ad oggi i dati provenienti dalla telemetria veicolare sono sfruttati nella pratica industriale principalmente per finalità operative quali diagnostica di bordo, manutenzione programmata e gestione delle flotte, senza essere integrati in un modello predittivo di valutazione dell'usura [2].

Attualmente, infatti, manca un approccio strutturato che, partendo dall'analisi storica dei dati telemetrici permetta di stimare in modo robusto il livello di usura dell'auto. Diventa quindi fondamentale sviluppare metodologie in grado di trasformare il patrimonio informativo generato dalla telemetria in indicatori utili nel processo di valutazione dell'auto, a supporto di un processo decisionale ancora fortemente dipendente dai controlli tradizionali, contribuendo a rendere più affidabile, trasparente e standardizzato l'intero processo di valutazione dell'usato [3][4][5].

L'adozione di strumenti analitici basati sui dati telemetrici rappresenterebbe un'evoluzione significativa che permetterebbe di superare i limiti delle attuali procedure di controllo, introducendo metriche quantitative, indicatori predittivi e modelli di apprendimento automatico in grado di stimare l'usura residua, il rischio di guasto e lo stato generale del veicolo [3][4][7].

1.3 Obiettivo della tesi

Nel contesto attuale di rapida innovazione del settore automotive e di crescente rilevanza strategica del mercato dell'usato garantito, il presente elaborato di tesi si propone di colmare il divario esistente tra la vasta disponibilità di dati telemetrici on-board e i tradizionali processi di valutazione dello stato di usura dei veicoli [1][3][4][7].

L'obiettivo dello studio è quello di esplorare come i dati grezzi raccolti dai sistemi telemetrici possano essere trasformati in conoscenza utile, sviluppando un approccio metodologico capace di stimare lo stato di salute del veicolo. Si vuole infatti dimostrare come un approccio data-driven possa superare i limiti delle metodologie attuali basate unicamente su controlli manuali, consentendo l'individuazione di pattern comportamentali, anomalie e segnali precoci di deterioramento non rilevabili attraverso un'ispezione convenzionale e valorizzando informazioni già disponibili ma ancora ampiamente sottoutilizzate.

In particolare, la ricerca mira a verificare la fattibilità e l'efficacia di un modello analitico che, attraverso l'analisi storica dei data-streams telematici e l'impiego di tecniche di Machine Learning, consenta di trasformare dati grezzi in indicatori diagnostici e predittivi affidabili. Più nello specifico il lavoro si propone di:

- identificare i sottosistemi e i parametri telemetrici più significativi per la valutazione dell'usura, con particolare attenzione al motore, sistemi di trasmissione e sistemi ausiliari del veicolo;
- analizzare lo storico dei dati di utilizzo al fine di ricostruire l'evoluzione temporale dei fenomeni di degradazione, tenendo conto delle condizioni operative e dello stile di guida;

- sviluppare un modello analitico basato su tecniche di Machine Learning capace di interpretare i dati raccolti dai sistemi di telemetria veicolare e stimare lo stato di usura del veicolo,
- stimare il potenziale impatto economico ed ambientale dell'adozione di tale strumento nel mercato dell'usato garantito come supporto ai processi di certificazione e rivendita dell'usato.

L'obiettivo ultimo è dunque mostrare come l'integrazione tra dati telemetrici e modelli predittivi possa costituire un valore aggiunto per l'intero ecosistema automotive, rendendo la valutazione dell'usato più affidabile, trasparente ed efficiente. Un approccio di questo tipo, infatti, riduce la dipendenza esclusiva dalle ispezioni manuali, valorizzando un patrimonio informativo esistente e standardizzando il processo valutativo, aprendo nuove opportunità per i costruttori e operatori del settore.

In questa prospettiva, la tesi non si limita a proporre un modello tecnico, ma intende evidenziare come la telemetria possa costituire un vero e proprio fattore abilitante per l'evoluzione dei programmi di usato certificato, favorendo l'adozione di processi analitici più oggettivi, riproducibili e basati su dati reali di utilizzo.

CAPITOLO 2

LA TELEMETRIA

2.1 Definizione e principi

Il termine *telemetria* deriva dal greco *tele* (“a distanza”) e *metron* (“misura”) ed indica l’insieme delle tecnologie, delle metodologie e dei dispositivi che consentono la misurazione, la trascrizione e la trasmissione a distanza di informazioni relative allo stato di un sistema fisico [9].

Nel dettaglio, in un sistema telemetrico i dati sono rilevati da sensori o da dispositivi remoti, digitalizzati tramite un’unità di acquisizione e successivamente trasmessi attraverso canali di telecomunicazione ad un centro di raccolta, un operatore o una piattaforma informatica dedicata all’analisi. Dal punto di vista strutturale la telemetria si basa su tre elementi fondamentali che ne definiscono l’architettura: la sorgente del dato, costituita da sensori, attuatori o centraline di controllo che misurano parametri fisici (temperatura, pressione, velocità, vibrazioni, tensioni elettriche, ecc.); il canale di trasmissione, che permette il trasferimento dei dati attraverso differenti tecnologie (radiofrequenza, sistemi satellitari, rete cellulare, Wi-Fi, Bluetooth o protocolli proprietari) e l’unità di destinazione, che riceve, memorizza, elabora e interpreta i dati per finalità di monitoraggio, diagnostica o previsione [3][4].

Dal punto di vista del flusso informativo, la comunicazione tra dispositivo telemetrico e centro di raccolta può avvenire secondo due modalità: monodirezionale o bidirezionale. La prima modalità di trasmissione consente l’invio di dati dai sensori verso la piattaforma centrale, senza la possibilità di intervento remoto da parte dell’operatore, rendendo questa prima configurazione particolarmente adatta ad applicazioni di monitoraggio passivo dove non è richiesto un controllo attivo. La trasmissione bidirezionale, al contrario, abilita un canale di comunicazione a doppio senso che, oltre alla ricezione dei dati, permette alla piattaforma centrale di inviare comandi operativi, aggiornamenti o richieste di configurazioni personalizzate al dispositivo, rendendo la telemetria un sistema integrato con duplice funzione di controllo e feedback [3][4][11].

Un ulteriore elemento distintivo riguarda la temporalità della trasmissione. I sistemi telemetrici possono operare in tempo reale, fornendo un flusso continuo di dati aggiornati, indispensabile per applicazioni critiche come controllo dinamico, sicurezza o monitoraggio di sistemi sensibili, oppure in modalità differita, in cui i dati vengono prima registrati localmente e poi inviati alla piattaforma di raccolta in un secondo momento, ad esempio per ottimizzare i consumi energetici, compensare limiti di connettività o gestire grandi quantità di informazioni [4][11][12][14].

2.2 Campi di applicazione

La telemetria è una tecnologia trasversale che trova applicazione in un'ampia varietà di settori grazie alla sua capacità di monitorare a distanza sistemi complessi, acquisire dati in modo continuo e supportare processi decisionali basati sulle informazioni raccolte. Negli ultimi decenni, lo sviluppo e la diffusione di sensori avanzati, reti di comunicazione e piattaforme di analisi dei dati ha ulteriormente accelerato la sua adozione, ampliandone in modo significativo le possibilità di utilizzo [3].

2.2.1 Monitoraggio remoto di sistemi complessi

Uno tra i primi e più consolidati utilizzi della telemetria riguarda il monitoraggio a distanza di sistemi complessi e difficilmente accessibili, come ad esempio veicoli spaziali, satelliti e sonde interplanetarie, per i quali la telemetria rappresenta l'unico mezzo per ricevere informazioni sui dati raccolti e sullo stato dei sistemi di bordo. L'utilizzo della telemetria interessa anche missili e dispositivi aeronautici, al fine di controllare parametri critici durante fasi operative delicate, inoltre essa trova anche applicazione in contesti come quelli di infrastrutture industriali ad alta criticità, come ad esempio piattaforme petrolifere offshore, impianti chimici, raffinerie e sistemi di generazione energetica, in quanto consente una migliore valutazione delle condizioni di sicurezza, permettendo di prevenire guasti e attivare tempestivamente le procedure correttive [2][12].

2.2.2 Ambito medico

La telemetria trova applicazione anche in ambito medico, nel quale svolge un ruolo fondamentale nel monitoraggio continuo dei parametri vitali in casi di anomalie cardiache o in pazienti che presentano condizioni potenzialmente critiche. La telemetria in ambito medico è

possibile attraverso dispositivi portatili o sensori indossabili che monitorano e registrano parametri come l'elettrocardiogramma (ECG), la frequenza cardiaca, i livelli di ossigenazione, la pressione arteriosa e altri fattori di interesse, consentendo al personale medico di analizzare i dati raccolti, facilitando diagnosi tempestive, terapie personalizzate e interventi proattivi [2][3].

2.2.3 Settore automotive

Il settore automotive è indubbiamente uno dei campi di applicazione in cui la telemetria è maggiormente utilizzata, spaziando dalla diagnostica remota alle applicazioni per il fleet management, fino ad innovazioni come gli Advanced Driver Assistance Systems (ADAS). Nel dettaglio, uno dei contesti storicamente più significativi per l'applicazione della telemetria automobilistica è quello delle competizioni, dove la telemetria rappresenta uno strumento imprescindibile per l'analisi delle prestazioni e del comportamento dinamico dei veicoli. In discipline come Formula 1, WRC ed endurance, le scuderie monitorano in tempo reale centinaia di parametri relativi a motore, aerodinamica e pneumatici, consentendo interventi immediati e decisioni basate su dati ad alta frequenza [2][3][4][11][13].

Oggi, l'utilizzo della telemetria in ambito automobilistico ha superato i confini delle gare automobilistiche, diffondendosi anche all'interno dei veicoli di uso quotidiano, consentendo al guidatore di avere un'esperienza di guida completa e moderna, che ne migliora non solo la sicurezza, ma anche l'efficienza energetica e il comfort [3][4].

2.2.4 Risparmio energetico

La telemetria è una tecnologia che sta trovando sempre più spazio anche in ambito energetico. In nazioni come ad esempio il Giappone, i sistemi telemetrici sono utilizzati come strumento per il monitoraggio dei consumi energetici di fabbriche, uffici e abitazioni, contribuendo in questo modo alla riduzione dello spreco di energia e al raggiungimento degli obiettivi ambientali, ottimizzando l'efficienza degli impianti e permettendo l'implementazione di politiche energetiche intelligenti [2][14].

2.2.5 Gestione e studio della fauna

La telemetria è impiegata anche in ambito ecologico e zoologico per il monitoraggio di animali selvatici e di specie a rischio di estinzione. Attraverso collari GPS, sensori biometrici e sistemi di tracciamento remoto, è possibile studiare gli spostamenti e i comportamenti degli animali, monitorare i loro parametri vitali, le loro interazioni con l'habitat naturale e rilevare situazioni

di pericolo o bracconaggio, supportando programmi di conservazione e gestione sostenibile della fauna [2].

2.2.6 Monitoraggio di macchinari industriali da remoto

Nel settore industriale, la telemetria consente il monitoraggio continuo di macchinari e impianti produttivi, rilevando dati relativi ai cicli di produzione, alla qualità del prodotto e alla comparsa di malfunzionamenti o anomalie, riducendo i tempi di fermo macchina e ottimizzando l'efficienza produttiva, rendendo possibile l'adozione di strategie di manutenzione predittiva [2][13][15].

2.3 Telemetria automobilistica

Il mercato della telemetria integrata nel settore automotive sta vivendo una fase di forte espansione, configurandosi come uno dei principali motori della trasformazione verso il veicolo connesso. Le analisi di settore stimano una crescita significativa del fatturato nel prossimo decennio con un CAGR superiore al 18% nella fascia temporale 2025–2035 e con un aumento dei ricavi da 23,71 miliardi di dollari del 2025 ai 130,55 miliardi di dollari stimati su scala mondiale per il 2035 [10].

Dal punto di vista geografico, il Nord America si conferma leader nell'adozione di soluzioni telematiche integrate, con una quota del mercato globale nel 2025 di oltre il 41,2% e un ruolo trainante nello sviluppo di standard e regolamentazioni per i veicoli connessi [4][10].

L'Europa mantiene una presenza significativa nel mercato grazie ad un solido settore automobilistico e ad una domanda crescente di tecnologie connesse spinta da interventi normativi che privilegiano sostenibilità e sicurezza. In questo contesto, Germania e Francia emergono come pilastri europei per la telemetria automobilistica, dove il primo integra i sistemi telemetrici nei veicoli per supportare i sistemi avanzati di assistenza alla guida (ADAS¹) e la diagnostica in tempo reale, grazie ad investimenti in R&S promossi dalla Verband der Automobilindustrie (VDA²) volti a migliorare la connettività e la sicurezza dei dati, rispondendo

¹ ADAS (Advanced Driver Assistance Systems) sono tecnologie di assistenza alla guida che utilizzano sensori, telecamere e centraline elettroniche per monitorare l'ambiente circostante e supportare il conducente. Tra le funzioni più comuni vi sono la frenata automatica d'emergenza, il mantenimento della corsia, il controllo della distanza di sicurezza e il rilevamento di distrazione o stanchezza, con l'obiettivo di aumentare la sicurezza e ridurre il rischio di incidenti.

² VDA (Verband der Automobilindustrie) associazione dell'industria automobilistica tedesca, organismo che rappresenta i principali costruttori e fornitori nazionali e che definisce standard tecnici e di qualità per il settore.

alla richiesta di veicoli più intelligenti e sicuri. La Francia, invece, attraverso iniziative come il piano “Francia 2030”³ promuove l’adozione di tecnologie per veicoli connessi evidenziando il ruolo della telematica negli obiettivi ambientali. Infatti, come riportato dall’Associazione dell’Industria Automobilistica Francese (CCFA), i consumatori francesi hanno mostrato la loro preferenza per funzionalità come monitoraggio salute veicolo e aggiornamenti traffico, consolidando il ruolo del Paese nella crescita del mercato della telemetria automobilistica europea [10][11][14].

Andando ad analizzare nel dettaglio le dinamiche del settore, la crescita è trainata da tre fattori principali: domanda di connettività veicolare e gestione flotte, aumento degli obblighi normativi e integrazione con guida autonoma. Questi driver influenzano il mercato non solo tramite l’impatto sul CAGR, ma anche attraverso altri parametri come l’influenza normativa (quanto quel driver è legato a leggi, regolamenti o obblighi imposti dalle autorità), la rilevanza geografica (in quali regioni ha maggiore impatto sul mercato e spillover⁴), il tasso di adozione (la velocità con cui si traduce in implementazioni concrete sui veicoli) e la cronologia dell’impatto (l’orizzonte temporale in cui produce effetti sulla crescita). Nella tabella seguente si analizzano gli effetti dei tre driver citati sui singoli parametri (Tabella 1) [4][10][11][12].

Tabella 1: Valutazione dei principali driver di crescita del settore in base a: impatto sul CAGR, influenza normativa, rilevanza geografica, tasso di adozione e cronologia dell’impatto

Parametro	Impatto CAGR	Influenza normativa	Rilevanza geografica	Tasso di adozione	Cronologia dell’impatto
Crescente domanda di soluzioni per la connettività veicolare e gestione flotte	3.80%	A breve termine (≤ 2 anni)	Asia Pacifico, Nord America (spillover: Europa)	Basso	Veloce
Aumento degli obblighi normativi	3.00%	Medio termine (2–5 anni)	Europa, Asia Pacifico (spillover: Nord America)	Alto	Moderare
Integrazione con sistemi di guida autonoma e ADAS	2.00%	A lungo termine (5+ anni)	Nord America, Europa	Medio	Lento

³ Francia 2030 è un piano nazionale di investimenti lanciato dal governo francese per rafforzare la sovranità industriale e tecnologica e sostenere la transizione ecologica entro il 2030, con fondi dedicati tra l’altro a mobilità elettrica, digitalizzazione e innovazione industriale.

⁴ spillover indicano gli effetti indiretti e positivi che un driver di crescita genera in regioni o mercati diversi da quelli di impatto principale, attraverso meccanismi di diffusione tecnologica, normativa o commerciale.

La forte crescita della telemetria ha spinto le case automobilistiche ad offrire un'ampia gamma di applicazioni che integrano questa tecnologia. Tra i servizi più diffusi si trovano le funzionalità di emergenza (eCall) e assistenza stradale, la localizzazione del veicolo e il tracciamento dei percorsi, la diagnostica remota con lettura dei codici di errore e monitoraggio dei parametri di funzionamento, nonché gli aggiornamenti software over-the-air (OTA⁵) delle centraline e dei sistemi di infotainment. A questi si aggiungono servizi orientati all'esperienza utente, come la navigazione con informazioni sul traffico in tempo reale, lo streaming di contenuti multimediali, la connettività Wi-Fi di bordo e l'integrazione con assistenti virtuali basati su intelligenza artificiale [4][13].

Un ulteriore ambito in forte crescita riguarda i servizi di comfort e di gestione remota del veicolo. Attraverso applicazioni dedicate è possibile controllare da remoto alcune funzioni, come il blocco e lo sblocco delle portiere, la pre-climatizzazione dell'abitacolo, la gestione della ricarica nei veicoli elettrici e la localizzazione dell'ultima posizione di parcheggio. In combinazione con i dati telemetrici, queste funzionalità consentono non solo di migliorare l'esperienza di utilizzo per il singolo conducente, ma anche di abilitare modelli di servizio evoluti (car sharing, flotte aziendali connesse) alla base dei nuovi paradigmi di mobilità intelligente [11][14].

2.4 Trasformazione dell'esperienza di guida

La diffusione della telemetria nel settore automobilistico non è dovuta soltanto a un'evoluzione tecnologica, ma riflette un cambiamento più profondo nella percezione stessa dell'automobile, che oggi non è più considerata un semplice mezzo di trasporto, ma bensì una vera e propria piattaforma di esperienze personalizzate e connesse [3][11].

Come visto nel paragrafo precedente, l'aspetto che impatta maggiormente sul tasso di crescita annuo (CAGR) è la domanda di connettività con un contributo di circa 3,8%, che mette in luce la crescente richiesta da parte dei consumatori di veicoli che offrano migliori prestazioni e maggiore sicurezza attraverso l'utilizzo di sistemi telemetria, capaci di raccogliere e analizzare i dati in tempo reale. Quest'aspetto offre da un lato nuove riflessioni in merito alle strategie di investimento delle case automobilistiche, con al centro l'integrazione dei sistemi telemetrici, ma

⁵ OTA è la capacità di un veicolo di scaricare e installare un nuovo software in modalità wireless, utilizzando una connessione a internet e senza la necessità di un collegamento fisico a un computer o di un intervento in concessionaria.

dall'altro, questo dato porta a riflettere in merito al cambiamento che sta avvenendo nella percezione del consumatore, che non guarda più all'auto come un mero mezzo di trasporto ma come una vera e propria esperienza di guida [11].

Grazie alla telemetria, l'auto si trasforma in un ambiente digitale e dinamico, capace di acquisire e analizzare in tempo reale dati sul comportamento del conducente, sulle condizioni di guida e sulle preferenze personali, adattando di conseguenza parametri come climatizzazione, illuminazione, posizioni sedili e suggerimenti di percorso. Questo approccio consente di offrire esperienze di guida uniche, caratterizzate da un livello sempre maggiore di comfort e personalizzazione. In questo senso, la telemetria segna il passaggio dell'auto da strumento funzionale ad un ecosistema esperienziale, diventando la base tecnologica utilizzata dai costruttori per rispondere alle nuove esigenze del mercato, sempre più orientate verso la ricerca di comfort e benessere alla guida. I dati telemetrici, integrati con algoritmi di machine learning, permettono di personalizzare l'interfaccia utente, prevedere le necessità del conducente e ottimizzare il consumo energetico in funzione dello stile di guida individuale [3][4][11].

Questo shift culturale e la continua evoluzione dell'industria automobilistica verso i veicoli a guida autonoma rende essenziale la presenza di sistemi telemetrici affidabili, capaci di supportare lo scambio di dati in tempo reale tra veicolo, infrastruttura e cloud computing. Questi flussi informativi, infatti, sono fondamentali per il processo decisionale dei sistemi autonomi, in quanto consentono agli algoritmi di intelligenza artificiale di elaborare scenari complessi come ostacoli imprevisti, condizioni meteorologiche variabili o interazioni con altri utenti della strada e di garantire livelli di sicurezza elevatissimi attraverso interventi predittivi e reattivi. Leader del settore come Tesla e Waymo sfruttano già questa sinergia analizzando i dati telemetrici raccolti durante la guida reale per affinare i modelli di intelligenza artificiale al fine di prevedere scenari complessi, incrementare la sicurezza dei veicoli e migliorare l'esperienza di guida dell'utente [12].

La combinazione di questi elementi apre nuove strade verso la collaborazione tra case automobilistiche e aziende tecnologiche, evolvendo il mercato della telemetria integrata e creando un panorama ricco di opportunità di innovazione e partnership strategiche. In definitiva, la telemetria sta ridefinendo l'esperienza di guida, rispondendo alle aspettative di una generazione di utenti per cui l'automobile è sinonimo di tecnologia, personalizzazione e qualità della vita in movimento.

2.5 Normative sulla sicurezza

Un altro driver a influenzare in misura determinante l'adozione dei sistemi di telemetria nel settore automobilistico sono le normative sulla sicurezza introdotte a livello internazionale ed europeo.

Gli enti regolatori hanno imposto standard tecnici sempre più avanzati con l'obiettivo di ridurre il numero di incidenti, migliorare la gestione delle emergenze e rendere le strade più sicure per tutti gli utenti. Questo quadro normativo ha trasformato la telemetria da tecnologia opzionale a componente strutturale dell'automobile moderna, accelerando la sua diffusione e definendo standard minimi che spingono l'intero settore verso un approccio data-driven orientato alla prevenzione [10].

Un caso emblematico è rappresentato dal sistema eCall, reso obbligatorio dal 31 marzo 2018 per tutti i nuovi modelli di autovetture e veicoli commerciali leggeri omologati nell'Unione Europea, nei Paesi EFTA⁶ e nel Regno Unito. Disciplinato dal Regolamento UE 2015/758, il sistema sfrutta tecnologie telematiche e di geolocalizzazione per attivare automaticamente una chiamata di emergenza al 112 in caso di grave incidente, trasmettendo dati essenziali come posizione del veicolo, direzione di marcia, timestamp dell'evento, tipologia del veicolo e altre informazioni utili ai soccorsi. L'introduzione dell'eCall rappresenta uno dei primi esempi concreti di obbligo normativo che ha reso la telemetria un requisito strutturale nei veicoli moderni, normalizzandone l'adozione su scala di massa [17].

A questo si aggiunge il Regolamento UE 2019/2144, che stabilisce l'obbligo per i costruttori di integrare nei veicoli una serie di sistemi avanzati di assistenza alla guida (ADAS), tra cui la frenata automatica di emergenza (AEB), il mantenimento della corsia (LKA) e il rilevamento della stanchezza del conducente. Tutti questi dispositivi richiedono la raccolta, l'elaborazione e, in alcuni casi, la trasmissione di dati telemetrici per monitorare continuamente l'ambiente circostante e il comportamento del veicolo, rendendo la telemetria un prerequisito tecnico indispensabile per adempiere agli obblighi regolatori [18]. Ulteriori normative internazionali, come i regolamenti UN R131 e UN R79 della Commissione Europea, prescrivono i sistemi di sicurezza attiva avanzata per veicoli, nel dettaglio l'Advanced Emergency Braking System

⁶ I Paesi EFTA (European Free Trade Association - Associazione Europea di Libero Scambio) comprendono attualmente Islanda, Liechtenstein, Norvegia e Svizzera.

(AEBS⁷) per R131 e i sistemi di sterzo assistito (ACSF⁸) integrati con Electronic Stability Control (ESC⁹) per R79, il cui funzionamento dipende da una costante acquisizione di dati telemetrici per interventi correttivi in tempo reale [19][20].

La telemetria è inoltre strettamente connessa alle normative ambientali, tra cui gli standard Euro 6 definiti dal Regolamento UE 2016/646, che stabiliscono limiti stringenti per le emissioni e le disposizioni del Regolamento UE 2024/1312, che ampliano i requisiti per il monitoraggio in tempo reale del corretto funzionamento dei sistemi antinquinamento [21][22].

Il mercato della telematica integrata deve altresì affrontare vincoli significativi derivanti da regolamenti sulla privacy dei dati, sicurezza informatica e sostenibilità ambientale, primo fra tutti il Regolamento generale sulla protezione dei dati (GDPR) dell'Unione Europea UE 2016/679, che impone linee guida rigorose sulla gestione dei dati e sul consenso degli utenti [23].

In conclusione, gli interventi normativi hanno spinto fortemente l'innovazione e l'adozione della telemetria veicolare, incentivando lo sviluppo di sistemi avanzati che rendessero non solo l'esperienza di guida più sicura attraverso interventi predittivi e correttivi, ma che pongono anche una particolare attenzione all'impatto ambientale, allineandosi agli obiettivi del Green Deal europeo per una mobilità sostenibile.

⁷ AEBS (Advanced Emergency Braking System) è il sistema di frenata d'emergenza avanzata che rileva automaticamente collisioni frontali imminenti tramite sensori e applica decelerazioni autonome.

⁸ ACSF (Advanced Control Steering Function) è una funzione avanzata di sterzo assistito, che include correzioni di traiettoria e Lane Keeping Assist, integrati con supervisione del conducente.

⁹ ESC (Electronic Stability Control) è il controllo elettronico della stabilità che previene sovrasterzo/sottosterzo tramite freni selettivi e gestione motore, basato su sensori yaw/roll.

CAPITOLO 3

SENSORISTICA DI BORDO

3.1 Diagnostica del veicolo

Spinte da imposizioni normative e dalla crescente domanda di veicoli sensorizzati, le innovazioni nel campo della telemetria automobilistica continuano a trasformare il paradigma della diagnostica di bordo. Oggi la telemetria automobilistica è la tecnologia alla base del sistema OBD (On-Board Diagnostics) capace di monitorare in tempo reale il funzionamento dei principali componenti del veicolo, rilevando anomalie, errori o guasti. Attraverso la presenza di sensori, le attuali funzioni della diagnostica di bordo sono in grado di raccogliere, elaborare e trasmettere in tempo reale una vasta gamma di parametri funzionali che spaziano dallo stato del motore alla temperatura dei liquidi, dai sistemi di sicurezza attiva al comportamento dinamico del veicolo, fino al controllo delle emissioni e del funzionamento dei dispositivi antinquinamento.

L'utilizzo della telemetria di bordo, infatti non migliora soltanto la sicurezza e il comfort durante la guida, ma rappresenta una vera e propria innovazione per quanto riguarda il controllo del funzionamento e delle prestazioni del veicolo, segnando il passaggio dalla manutenzione programmata alla manutenzione predittiva. Grazie all'integrazione della telemetria all'interno della vettura, la diagnostica non è più un'attività eseguita solo in occasione della revisione o dell'ingresso in officina, ma diventa un processo continuo e proattivo, in grado di individuare tempestivamente anomalie o deterioramenti prestazionali prima che si trasformino in guasti veri e propri. L'analisi continua di dati telemetrici consente quindi di migliorare l'affidabilità dei veicoli, ridurre i tempi di fermo, ottimizzare gli interventi manutentivi e garantire livelli più elevati di sicurezza ed efficienza operativa [3][4].

In questo nuovo paradigma, il veicolo diventa un nodo intelligente all'interno di una rete più ampia che, partendo dai sensori integrati all'interno della vettura, arriva alle concessionarie per la diagnostica remota e le pianificazioni di interventi preventivi.

3.2 Architettura telemetria automobilistica

Entrando nel dettaglio dell'architettura, i moderni sistemi di telemetria automobilistica si basano su tre elementi principali: sensori e centraline (ECU), reti di comunicazione e piattaforme di raccolta ed elaborazione dati.

Per quanto riguarda i sensori, questi sono dispositivi di misurazione di input che hanno il compito di rilevare e misurare una specifica grandezza fisica o chimica nell'ambiente o in un componente del veicolo. Il parametro di interesse, dopo essere stato rilevato dal sensore, viene convertito in un segnale elettrico (analogico o digitale) in modo da poter essere letto ed elaborato dalla centralina ad esso associata. Le centraline ECU (Electronic Control Unit) sono microprocessori di bordo fondamentali per gestire attivamente i sottosistemi del veicolo. Questi, infatti, sono dispositivi di controllo e decisione che ricevono in tempo reale i data-streams operativi provenienti dai sensori (funzione di acquisizione dati) e li elaborano istantaneamente per generare una risposta di regolazione (funzione decisionale) che viene condivisa con il resto del veicolo tramite un protocollo di rete digitale (come ad esempio il CAN-Bus), garantendo l'ottimizzazione immediata delle performance e del controllo dei sistemi [11].

Le reti di comunicazione sono essenziali per consentire la comunicazione e la trasmissione dei dati tra le diverse centraline, con i sensori ad esse connesse e con le componenti associate per la regolazione delle funzioni del veicolo, creando una vera e propria rete di bordo (*In-Vehicle Network*). Nel dettaglio, esistono diversi protocolli che possono essere utilizzati per la comunicazione che si differenziano tra di loro per velocità di trasmissione, affidabilità e costo, e pertanto impiegati in contesti differenti. Nella tabella si analizzano i principali protocolli utilizzati all'interno delle automobili (Tabella 2) [11].

Tabella 2: Confronto tra i principali protocolli presenti nelle reti automobilistiche per utilizzo, velocità di trasmissione e caratteristiche principali.

Protocollo	Utilizzo Tipico (Funzione)	Velocità di Trasmissione	Caratteristiche Principali
CAN-Bus (Controller Area Network)	Standard universale per il <i>powertrain</i> (motore, trasmissione), freni (ABS) e airbag.	Bassa/Media (fino a 1Mbit/s)	Robustezza e affidabilità elevate. Ottimo per dati critici in tempo reale.
LIN (Local Interconnect Network)	Applicazioni a bassa velocità e non critiche (sensori semplici, comandi dei sedili, finestrini, climatizzatore).	Molto Bassa (fino a 20kbit/s)	Basso costo e architettura semplice (un solo master e nodi slave).

FlexRay	Sistemi che richiedono altissima affidabilità e sincronizzazione temporale (sistemi <i>drive-by-wire</i> , sospensioni attive).	Alta (fino a 10Mbit/s)	Comunicazione deterministica (garantisce i tempi di consegna), tolleranza ai guasti.
Automotive Ethernet	Infotainment (video HD), diagnostica ad alta velocità e ADAS (guida autonoma).	Molto Alta (fino a 1Gbit/s)	Velocità elevatissima, scalabilità e uso di <i>standard IT</i> noti.

Un ulteriore elemento essenziale per descrivere l'architettura della telemetria automobilistica sono le piattaforme di raccolta ed elaborazione dei dati, le quali hanno il compito di trasformare il flusso di dati grezzi provenienti dal veicolo in informazioni strutturate. Nel dettaglio, le piattaforme di raccolta comprendono i moduli di acquisizione, che ricevono e validano i dati trasmessi dalle unità telematiche di bordo, i sistemi di memorizzazione, che consentono l'archiviazione storica delle misure e i motori di elaborazione, che applicano algoritmi di filtraggio, aggregazione e analisi avanzata. I risultati di tali elaborazioni vengono infine resi disponibili ai diversi attori coinvolti (costruttori, gestori di flotte, officine, conducenti), abilitando servizi come la diagnostica remota, la manutenzione predittiva e l'ottimizzazione operativa [11].

3.3 Sottosistemi e parametri monitorati

Per sviluppare un modello predittivo capace di stimare lo stato di usura del veicolo a partire dall'utilizzo dei dati telemetrici disponibili, è necessario fare un'analisi di quelli che sono attualmente i parametri monitorati all'interno di un veicolo. A questo scopo si approfondiscono di seguito i principali i sottosistemi del veicolo che vengono monitorati tramite l'integrazione di sistemi telemetrici.

Essendo la ricerca mirata a supportare processi decisionali nella rivendita dell'usato garantito, si è scelto di focalizzare l'analisi sui sensori telemetrici standard presenti nei veicoli di serie post-2001 UE/1996 USA, anno di introduzione obbligatoria dell'EOBD (European On-Board Diagnostics). Tale delimitazione garantisce di avere dati uniformi e compatibili tra le diverse

case automobilistiche grazie all'introduzione dello standard CAN-BUS e di codici PID¹⁰ universali [3][24].

3.3.1 Motore e sistema di alimentazione

Il gruppo propulsore rappresenta uno dei sistemi più monitorati tramite interfacce OBD-II e costituisce una fonte primaria per valutare lo stato meccanico complessivo del veicolo. Di seguito si descrivono i principali parametri monitorati tramite PID OBD-II e la loro rilevanza in termini di stima dell'usura del veicolo: [25][26]

- Giri motore (RPM): Indicano quante rotazioni al minuto compie l'albero motore e descrivono il regime di funzionamento (minimo, parziale, pieno carico). Combinati con altri parametri, questi dati permettono di stimare stress meccanico e condizioni d'uso del propulsore.
- Carico motore (%): Misura in termini percentuali quanto il motore sta lavorando rispetto al suo potenziale massimo. Valori medi elevati su lunghi periodi indicano un utilizzo gravoso (es. traino, percorsi montani, guida sportiva), con potenziale accelerazione dei fenomeni di usura.
- Temperatura liquido di raffreddamento: Misura la temperatura del refrigerante che circola nel blocco motore. Valori a regime (intorno a 90 °C) indicano funzionamento corretto, mentre temperature cronicamente alte o forti oscillazioni possono segnalare problemi al circuito di raffreddamento o condizioni d'uso severe.
- Pressione del collettore (MAP)/Portata massa aria (MAF): il sensore MAP misura la pressione assoluta nel collettore di aspirazione ed è un indicatore del carico motore e della quantità d'aria che entra nei cilindri. Il sensore MAF misura direttamente la massa d'aria aspirata (ad esempio in g/s) ed è il riferimento principale per calcolare quanta benzina/gasolio iniettare per ottenere una combustione corretta. Insieme questi due parametri descrivono quanta aria "respira" il motore e sono quindi fondamentali per la stima dell'efficienza di riempimento e del corretto dosaggio del carburante, valori ridotti segnalano la perdita di compressione o restrizione dell'aspirazione, che può portare all'usura del motore.

¹⁰ I PID (Parameter ID) sono codici esadecimale standardizzati, definiti dal protocollo OBD-II/EOBD che permettono la richiesta dei parametri veicolari alla centralina ECU tramite la porta diagnostica.

- Consumo specifico carburante/lambda (sonde O₂): Il consumo specifico carburante è un parametro che misura la quantità di carburante necessaria per produrre un'unità di lavoro utile (1kWh di energia meccanica erogata dal motore). Questo parametro, noto anche come lambda (λ) è il rapporto normalizzato tra il consumo di carburante e la potenza termica prodotta, fornendo un indicatore compatto dell'efficienza termica complessiva del motore, indipendente da regime operativo, carico o condizioni d'uso. Questo è inoltre un parametro sintetico utile per descrivere l'efficienza del motore in modo compatto e universale, eliminando la dipendenza da variabili specifiche e permettendo confronti rapidi tra motori, esprimendo il rapporto tra aria effettiva e aria "ideale" nella miscela aria-carburante:

$\lambda = 1 \rightarrow$ miscela stechiometrica (combustione ottimale)

$\lambda > 1 \rightarrow$ miscela magra

$\lambda < 1 \rightarrow$ miscela grassa

Le sonde lambda misurano l'ossigeno residuo nei gas di scarico e permettono alla centralina di correggere in tempo reale l'iniezione, mantenendo sotto controllo consumi ed emissioni e riducendo il rischio di usura.

- Anticipo accensione (spark advance): Indica di quanti gradi prima del Punto Morto Superiore (BTDC) viene fatta scoccare la scintilla alla candela. Valori sistematicamente troppo elevati o troppo ridotti possono indicare problemi di fasatura, anomalie nei sensori o una combustione non ottimale.
- Misfire sui cilindri: Sono gli "scoppi mancati", cioè eventi in cui in un cilindro non avviene una combustione corretta quando dovrebbe. Un numero elevato o ripetuto di misfire su uno o più cilindri segnala criticità come candele usurate, iniettori ostruiti, problemi di accensione o compressione, rappresentando un indicatore diretto del degrado del sistema di alimentazione/accensione.

I parametri analizzati sono stati riassunti nella tabella seguente riportando per ognuno il PID corrispondente, l'aspetto che si intende monitorare e l'unità di misura (Tabella 3).

Tabella 3: Per ciascun parametro del motore sono riportati i PID di riferimento, l'unità di misura e l'aspetto che si vuole verificare.

PID	Parametro	Controllo	Unità
010C	Giri motore	Motore stressato	Giri/min
0104	Carico motore	Carico gravoso	%

0105	Temperatura liquido raffreddamento	Motore surriscaldato	°C
010B	Pressione collettore	Carico aspirazione	kPa
0121	Lambda	Miscela aria-carburante	rapporto λ
010E	Anticipo accensione	Combustione non ottimale	gradi BTDC
P0308	Misfire cilindri	Scoppi mancati	conteggio DTC

3.3.2 Sistema di trasmissione e frizione

Il sistema di trasmissione subisce usura progressiva, particolarmente marcata nei veicoli ad alto chilometraggio o con utilizzo gravoso e rilevante per determinare il livello di usura complessivo dell'auto. Di seguito si descrivono i principali parametri monitorati e la loro rilevanza in termini di stima dell'usura del veicolo: [25][26]

- Temperatura olio cambio: I sensori NNTC o termocoppie resistenti misurano il calore del fluido (range tipico 80-260°F) in modo da prevenire degradazione e guasti da surriscaldamento.
- Velocità input/output: I sensori su ingressi e uscite del cambio rilevano posizione e velocità per regolare pressioni. Sincronizzano frizione e cambiate fluide, evitando malfunzionamenti.
- Pressione olio/idraulica: Sensori monitorano livelli pressione per attuatori e frizione. Prevedono slittamenti eccessivi e garantiscono cambi precisi.

I parametri analizzati sono stati riassunti nella tabella seguente riportando per ognuno l'aspetto che si intende monitorare e l'unità di misura. Si sottolinea che i codici per il rilevamento del valore nel caso dei sensori analizzati non sono PID OBD-II universali, ma possono differire per casa automobilistica (Tabella 4).

Tabella 4: Per ciascun parametro del sistema di trasmissione sono riportati l'unità di misura e l'aspetto che si vuole verificare.

Parametro	Controllo	Unità
Temperatura olio cambio	Surriscaldamento fluido	°C
Velocità input shaft	Slittamento frizione ingresso	giri/min
Velocità output shaft	Slittamento frizione uscita	giri/min
Pressione olio/idraulica	Perdite circuiti attuatori	kPa

3.3.3 Sistema di frenata

Il sistema frenante è tra i principali indicatori di sicurezza veicolare e rappresenta per tale motivo uno dei sottosistemi maggiormente sensorizzati. Di seguito si descrivono i principali parametri monitorati e la loro rilevanza in termini di stima dell'usura del veicolo: [25][26]

- Pressione circuito frenante (pre/post ABS): I sensori idraulici misurano la pressione in tempo reale, fornendo all'ECU i dati per modulazioni ABS che prevengono il blocco ruote. Tali dati sono inoltre utili per determinare l'usura delle pastiglie e dei dischi, rilevati per cali di pressione maggiori del 15%.
- Interventi ABS/ESP: L'ECU memorizza eventi di attivazione correlati a velocità ruote e accelerometri, identificando pattern di instabilità rilevanti per analisi di sicurezza e per determinare rischi di usura asimmetrica delle pastiglie.
- Livello fluido freni: Monitora cali di livello del liquido freni per rilevare perdite o bolle d'aria nei circuiti. I dati rilevati dai sensori sono utili per evitare guasti al sistema frenante e per determinare l'usura delle pastiglie.

I parametri analizzati sono stati riassunti nella tabella seguente riportando per ognuno l'aspetto che si intende monitorare e l'unità di misura. Si sottolinea che i codici per il rilevamento del valore nel caso dei sensori analizzati non sono PID OBD-II universali, ma possono differire per casa automobilistica (Tabella 5).

Tabella 5: Per ciascun parametro del sistema frenante sono riportati l'unità di misura e l'aspetto che si vuole verificare.

Parametro	Controllo	Unità
Pressione circuito frenante	Modulazioni ABS/integrità	bar/kPa
Interventi ABS/ESP	Instabilità/usura frenata	numero eventi
Livello fluido freni	Perdite/bolle circuito	MI

3.3.4 Pneumatici e dinamica del veicolo

La telemetria consente di monitorare diversi aspetti della dinamica veicolare rilevanti per verificare le condizioni operative del veicolo. Di seguito si descrivono i principali parametri monitorati e la loro rilevanza in termini di stima dell'usura del veicolo: [25][26]

- Velocità ruota (wheel speed): I sensori hall attivi o passivi su ogni ruota misurano i giri per calcolare la velocità differenziale tra ruote. Tali dati servono per ABS (antiblocco freni) ed ESP (controllo stabilità), rilevando squilibri o guasti.

- Pressione pneumatici (TPMS): I sensori diretti (valvolari) o indiretti (da wheel speed) controllano il gonfiaggio dello pneumatico, segnalando cali oltre il 20%. Obbligatori in UE e USA per sicurezza, riducono usura e rischi.
- Interventi ESP/traction control: L'ECU registra eventi di slittamento ruota e correzioni (frenata selettiva o riduzione potenza). Questi indicatori segnalano una guida sportiva o instabilità e sono utili per segnalare rischi di usura accelerata degli pneumatici.

I parametri analizzati sono stati riassunti nella tabella seguente riportando per ognuno l'aspetto che si intende monitorare e l'unità di misura. Si sottolinea che i codici per il rilevamento del valore nel caso dei sensori analizzati non sono PID OBD-II universali, ma possono differire per casa automobilistica (Tabella 6).

Tabella 6: Per ciascun parametro relativo agli pneumatici e dinamica del veicolo sono riportati l'unità di misura e l'aspetto che si vuole verificare.

Parametro	Controllo	Unità
Velocità ruota	Bloccaggio ruote	giri/min
Pressione pneumatici	Sgonfiaggio	bar/psi
Interventi ESP/traction control	Guida aggressiva/instabilità	Eventi

3.3.5 Sistema elettrico ed elettronico

Il sistema elettrico ed elettronico all'interno di un'auto fornisce informazioni fondamentali sullo stato della batteria e dei sistemi elettronici tramite sensori BMS/IBS e centraline su CAN/OBD-II. Di seguito si descrivono i principali parametri monitorati e la loro rilevanza in termini di stima dell'usura del veicolo: [25][26]

- Tensione di batteria (12 V): Parametro misurato tramite divisori resistivi collegati agli ingressi delle centraline o da sensori IBS¹¹ dedicati, permette di determinare lo stato di carica, la salute della batteria e il funzionamento del circuito di ricarica alternatore-regolatore. Tale parametro è monitorato attraverso standard OBD-II con PID 01 42.
- Corrente assorbita/erogata: Misurata dall'IBS tramite resistore shunt o sensore hall, questo parametro monitora il flusso elettrico in ingresso/uscita dalla batteria, rilevando inefficienze di ricarica e consumi anomali.

¹¹ IBS (Intelligent Battery Sensor, o Sensore Batteria Intelligente) è un modulo elettronico dedicato montato sul polo negativo della batteria ausiliaria a 12 V nelle auto a combustione interna moderne. Funziona come un "supervisore" della batteria, misurando parametri chiave e stimando il suo stato.

- Temperatura batteria: Rilevata tramite termistori NTC integrati nell'IBS, questo parametro rileva anomalie termiche del sistema elettrico, surriscaldamenti celle/moduli per garantire sicurezza e longevità della batteria.
- Resistenza interna/impedenza: Valuta lo stato interno della batteria misurando la “difficoltà interna” con cui la corrente elettrica fluisce attraverso le celle chimiche della batteria. Tale parametro restituisce un'importante informazione in merito alla bontà della batteria del veicolo, fornendo un elemento utile alla stima del suo livello di usura.
- Stato CAN BUS: Misura l'integrità della rete di comunicazione digitale tra tutte le centraline dell'auto, restituendo un parametro utile per determinare il livello di usura della rete del veicolo.
- DTC elettrici: Rilevano i codici di errore specifici registrati dalle centraline relative al sistema di alimentazione elettrica e rilevante per determinare il livello di deterioramento delle stesse.

I parametri analizzati sono stati riassunti nella tabella seguente riportando per ognuno l'aspetto che si intende monitorare e l'unità di misura. Si sottolinea che i codici per il rilevamento del valore nel caso dei sensori analizzati non sono PID OBD-II universali ma possono differire per casa automobilistica, ad eccezione del parametro tensione batteria per il quale è già stato riportato il codice PID relativo (Tabella 7).

Tabella 7: Per ciascun parametro del sistema elettrico ed elettronico sono riportati le unità di misura e l'aspetto che si vuole verificare.

Parametro	Controllo	Unità
Tensione batteria	Alimentazione ECU	V
Corrente assorbita/erogata	Flusso energia batteria	A
Temperatura batteria	Celle batteria	°C
Resistenza interna/impedenza	Degradazione interna batterie	mΩ
Stato CAN bus	Comunicazione centraline	Codici errore
DTC elettrici	Guasti sistema alimentazione	Codici errore

3.2.6 Sospensioni e telaio

Monitorare alcuni aspetti relativi alle sospensioni e telaio può essere utile per comprendere l'usura degli ammortizzatori e l'impatto delle condizioni di guida sul degrado strutturale del veicolo. Di seguito si descrivono i principali parametri monitorati e la loro rilevanza in termini di stima dell'usura del veicolo: [25][26]

- **Accelerazioni verticali:** I sensori IMU¹² misurano la forza che spinge l'auto lungo l'asse z (verso l'alto o il basso) in presenza di dossi, buche o asperità stradali. Un'accelerazione verticale di 1g corrisponde al peso normale dell'auto, mentre valori superiori indicano sollecitazioni anomale e usura delle sospensioni.
- **Accelerazioni laterali:** I sensori IMU rilevano la forza centrifuga che fa pendere l'auto lungo l'asse y (verso destra o sinistra) quando affronta una curva e sono rilevanti per determinare il livello di usura delle sospensioni.

I parametri analizzati sono stati riassunti nella tabella seguente riportando per ognuno l'aspetto che si intende monitorare e l'unità di misura. Si sottolinea che i codici per il rilevamento del valore nel caso dei sensori analizzati non sono PID OBD-II universali, ma possono differire per casa automobilistica (Tabella 8).

Tabella 8: Per ciascun parametro di sospensioni e telaio monitorato con relative unità di misura e l'aspetto che si vuole verificare.

PID	Parametro	Controllo	Unità
Proprietario	Accelerazioni verticali	Rimbalzo verticale	g
Proprietario	Accelerazioni laterali	Rollio curve	g

3.2.7 Sistema di scarico e controllo emissioni

L'analisi dei parametri relativi al sistema di scarico e al controllo delle emissioni è rilevante per determinare il livello di degradazione, intasamento o ostruzioni dei componenti, come richiesto dalle normative Euro 6 per garantire la conformità delle emissioni [21]. Di seguito si descrivono i principali parametri monitorati e la loro rilevanza in termini di stima dell'usura del veicolo: [25][26]

- **Temperatura gas di scarico:** Rileva la temperatura effettiva dei gas di scarico che fuoriescono dal motore del veicolo. Tale parametro viene misurato tramite sensori EGTS (Exhaust Gas Temperature Sensor) a termistore NTC integrato nel tubo di scarico, con valori che rientrano tipicamente nel range (200-900°C). Temperature anomale segnalano guasti specifici che possono indicare surriscaldamento del catalizzatore o mancato light-off, prevedendo la sostituzione del componente. Tali

¹² IMU (Inertial Measurement Unit) è un'unità di misura composta da accelerometro MEMS 3-assi e giroscopio che viene utilizzata per misurare accelerazioni lineari e rotazioni angolari per controllare la stabilità dinamica del veicolo.

parametri vengono rilevati attraverso due sensori differenti con PID standard OBD-II che sono rispettivamente 01 3C e 01 3D.

- **Efficienza catalizzatore:** Misura la capacità residua del catalizzatore di convertire sostanze inquinanti come idrocarburi (HC), monossido di carbonio (CO) e ossidi di azoto (NO_x), rispettivamente in anidride carbonica (CO₂), vapore acqueo (H₂O) e azoto molecolare (N₂), riducendo le emissioni del 90-99%. L'efficienza del catalizzatore viene misurata confrontando le oscillazioni lambda pre e post catalizzatore tramite sensori OBD-II. Tali parametri vengono rilevati attraverso due sensori differenti con PID standard OBD-II che sono rispettivamente 01 05 e 01 07.
- **Valvola EGR:** La valvola EGR (Exhaust Gas Recirculation) riposta indietro una parte dei gas di scarico nei cilindri per abbassare la temperatura di combustione e ridurre gli ossidi di azoto (NO_x). La funzionalità della valvola EGR è monitorata dal controllo ad anello chiuso che verifica la corretta apertura e monitora la massa di aria aspirata, indicando in caso di malfunzionamento usura da accumulo di carboniosi.
- **Filtro antiparticolato (DPF):** Il filtro antiparticolato DPF (Diesel Particulate Filter) cattura il 99% delle particelle di carbonio prodotte dal motore diesel e le brucia trasformandole in CO₂. I sensori monitorano il corretto funzionamento del filtro misurando il livello di intasamento tramite pressione differenziale.

I parametri analizzati sono stati riassunti nella tabella seguente riportando per ognuno l'aspetto che si intende monitorare e l'unità di misura. Si precisa che i PID OBD-II riportati per temperatura gas di scarico ed efficienza catalizzatore sono codici universali (SAE J1979), mentre valvola EGR e DPF utilizzano codici proprietari specifici per casa automobilistica (Tabella 9).

Tabella 9: Per ciascun parametro del sistema di scarico e controllo emissioni si riportano le unità di misura e l'aspetto che si vuole verificare.

Parametro	Controllo	Unità
Temperatura gas di scarico	EGTS pre/post catalizzatore	°C
Efficienza catalizzatore	Sonde lambda	%
Valvola EGR	Deviazione posizione valvola	%
Filtro antiparticolato (DPF)	Quantità fuliggine	G

L'analisi condotta ha permesso di identificare quali sono attualmente i sottosistemi dell'auto maggiormente monitorati tramite l'utilizzo della telemetria e quali le variabili di interesse per ciascuno di essi. Nel dettaglio, dallo studio emerge come particolare attenzione sia riservata al motore, che risulta il sottosistema maggiormente sensorizzato e, di conseguenza, quello con il

maggior numero di parametri monitorati in tempo reale, aspetto che sottolinea l'importanza cruciale ricoperta da questo componente nella valutazione delle prestazioni complessive del veicolo. Tale centralità è inoltre sottolineata dal fatto che, come emerso dallo studio, il motore rappresenta il sottosistema con il maggior numero di sensori associati a codici PID OBD-II universali, come stabilito dalle normative ISO 15765-2¹³ e SAE J1979¹⁴ [27][28]. La presenza di sensori all'interno del motore delle auto, quindi, non è soltanto una scelta legata alla rilevanza tecnica di questo componente al fine di monitorare le prestazioni ed analizzare il comportamento del veicolo, ma è anche imposta da normative, aspetti che rendono il motore il principale oggetto della telemetria automobilistica.

¹³ ISO 15765-2 definisce il protocollo di trasporto CAN per comunicazioni diagnostiche OBD-II (segmentazione, flow control). Standard obbligatorio per veicoli post-2008.

¹⁴ SAE J1979 (OBD-II) definisce le modalità di comunicazione tra il sistema di diagnostica a bordo (OBD) di un veicolo e gli strumenti di diagnostica esterni, come i lettori di codici di errore DTC.

CAPITOLO 4

ANALISI DEL DATASET

4.1 Selezione del dataset

Noti quali sono i principali sottosistemi sensorizzati e quali i parametri monitorati attraverso l'utilizzo della telemetria veicolare, per procedere con lo studio previsto nella presente tesi e sviluppare un modello predittivo di machine learning che a partire dai dati telemetrici sia in grado di stimare lo stato di usura del veicolo, risulta indispensabile avvalersi di un dataset contenente dati telemetrici raccolti a partire dai sistemi di bordo del veicolo.

A tale scopo, si è condotto uno screening sistematico e rigoroso dei dataset pubblici disponibili online, prioritizzando la ricerca su repository digitali affidabili e consolidati a livello accademico e scientifico quali Kaggle e RADAR (accessibile tramite re3Data.org). La selezione del dataset è stata inoltre orientata verso fonti caratterizzate da dimensioni adeguate (superiori a 1000 osservazioni), conformi allo standard OBD-II con parametri codificati mediante PID standardizzati secondo le normative ISO 15765-2 e SAE J1979 e aventi licenze open access compatibili con l'utilizzo in ambito di ricerca accademica [27][28]. Tali restrizioni nella ricerca sono state applicate allo scopo di trovare un dataset rappresentativo ed universale, in modo che il modello di machine learning sviluppato a partire dall'utilizzo dello stesso fosse robusto e applicabile su tutti i modelli automobilistici dotati di sensori telemetrici.

La scelta del dataset specifico ha altresì tenuto conto della necessità che esso includesse un insieme completo di grandezze fisiche tipiche della telemetria di serie installata sulle vetture moderne a combustione interna, come il regime motore (ENGINE_RPM), la temperatura del liquido di raffreddamento (COOLANT_TEMP), la posizione della farfalla del gas (THROTTLE_POS), il carico motore (ENGINE_LOAD), e parametri correlati al sistema di alimentazione (pressione carburante e flusso aria). Tali grandezze risultano infatti disponibili in modo diffuso e standardizzato sulle automobili contemporanee, garantendo una piena coerenza con il contesto applicativo della tesi e assicurando che i risultati dell'analisi derivante siano il più possibile universali e replicabili su un'ampia gamma di veicoli.

4.2 Descrizione del dataset

Coerentemente con i criteri di selezione del dataset illustrati, per lo sviluppo degli algoritmi di stima dello stato di usura del veicolo è stato utilizzato un dataset pubblico di telemetria veicolare, denominato *OBD2Data*, disponibile sulla piattaforma Kaggle, che raccoglie misure telemetriche acquisite tramite interfaccia diagnostica standard OBD-II su veicoli a combustione interna [29]. Il dataset è organizzato in formato tabellare (file CSV) e contiene, per ciascun campione temporale, un insieme di segnali standard OBD-II, tra cui regime motore, velocità veicolo, temperatura del liquido di raffreddamento, carico motore e parametri relativi al sistema di alimentazione. Queste grandezze risultano rappresentative delle sollecitazioni meccaniche e termiche cui il veicolo è sottoposto e sono tipicamente disponibili anche sulle vetture di serie tramite comuni dispositivi di lettura OBD-II, rendendo il dataset coerente con il contesto applicativo della presente tesi.

Il dataset raccoglie oltre 1,5 milioni di misure telemetriche relative a 27 parametri OBD-II standard (PID) monitorati su un veicolo Toyota Etios 2014 (cilindrata 1496 cm³) in condizioni di guida reale per una durata di 10 minuti ciascuna. L'acquisizione dei dati è stata effettuata mediante il dispositivo Carloop (kit open-source con connettività cellulare e interfaccia OBD-II) collegato alla porta diagnostica del veicolo e le misurazioni sono state effettuate con frequenza di 1 Hz (1 campione al secondo).

All'interno del dataset selezionato i dati risultano organizzati in cinque file, riconducibili alle seguenti condizioni di guida: *driveX.csv* (guida su strade ad alta velocità), *liveX.csv* (tragitto casa-lavoro), *ufpeX.csv* (guida all'interno del campus universitario a bassa velocità), *idleX.csv* (motore acceso a veicolo fermo) e *longX.csv* (viaggi lunghi).

Tra i diversi parametri monitorati emergono grandezze chiave per l'analisi di usura come: regime motore (*ENGINE_RPM*), carico calcolato (*ENGINE_LOAD*), temperatura del liquido di raffreddamento (*COOLANT_TEMPERATURE*), tensione del modulo di controllo (*CONTROL_MODULE_VOLTAGE*), posizione del pedale acceleratore (*ACCELERATOR_PEDAL_POSITION_D/E*), portata aria (*MAF* implicita nei parametri correlati) e parametri diagnostici come i codici di guasto (*TIME_RUN_WITH_MIL_ON*).

I parametri presenti nel dataset risultano tutti conformi allo standard SAE J1979 e, sebbene il dataset originale fosse orientato all'analisi dei pattern di comportamento del guidatore, nella presente tesi tali parametri saranno analizzati attraverso algoritmi di machine learning per

restituire una classificazione dello stato di usura del veicolo, costituendo in tal modo un indicatore sintetico che possa essere utilizzato per aiutare a stimare in modo oggettivo il livello di usura dei veicoli rivenduti sul mercato dell'usato garantito.

Di seguito è riportato l'elenco completo dei 27 parametri OBD-II contenuti all'interno del dataset *OBD2Data* selezionato per lo studio, riportando per ciascuno di essi il nome della variabile monitorata, l'unità di misura, la descrizione tecnica e la sua rilevanza nell'analisi di usura veicolare (Tabella 10) [29].

Tabella 10: Per ciascun parametro presente nel dataset selezionato si illustrano il nome della variabile, l'unità di misura, la descrizione tecnica e la rilevanza per la stima dell'usura veicolare.

Parametro	Unità	Descrizione	Usura
ENGINE_RUN_TIME	S	Tempo di funzionamento motore dall'ultimo avvio	Ore motore totali (usura cumulativa)
ENGINE_RPM	Rpm	Regime motore (giri al minuto)	Stress meccanico ad alti regimi
VEHICLE_SPEED	km/h	Velocità veicolo	Profilo di utilizzo (urbano/extra urbano)
THROTTLE	%	Posizione farfalla gas	Stile di guida aggressivo
ENGINE_LOAD	%	Carico motore calcolato dall'ECU	Sollecitazioni elevate motore
COOLANT_TEMPERATURE	°C	Temperatura liquido raffreddamento	Stress termico componenti motore
LONG_TERM_FUEL_TRIM_BANK_1	%	Correzione miscela lungo termine	Efficienza combustione/ usura iniettori
SHORT_TERM_FUEL_TRIM_BANK_1	%	Correzione miscela breve termine	Regolazioni dinamiche miscela
INTAKE_MANIFOLD_PRESSURE	kPa	Pressione assoluta collettore aspirazione	Carico motore e turbolenza
FUEL_TANK_LEVEL_INPUT	%	Livello carburante serbatoio	Consumi e stile di guida
ABSOLUTE_THROTTLE_B	%	Posizione assoluta farfalla	Controllo motore

ACCELERATOR_PEDAL_POSITION_D	%	Posizione pedale acceleratore	Intensità accelerazioni
ACCELERATOR_PEDAL_POSITION_E	%	Posizione pedale acceleratore	Intensità accelerazioni (ridondante)
COMMANDED_THROTTLE_ACTUATOR	%	Attuatore farfalla comandato	Risposta sistema controllo
FUEL_AIR_COMMANDED_EQUIV_RATIO	Ratio	Rapporto stechiometrico comandato	Efficienza combustione
ABSOLUTE_BAROMETRIC_PRESSURE	kPa	Pressione barometrica assoluta	Condizioni ambientali operative
RELATIVE_THROTTLE_POSITION	%	Posizione relativa farfalla	Dinamica acceleratore
INTAKE_AIR_TEMPERATURE	°C	Temperatura aria in ingresso motore	Condizioni termiche aspirazione
TIMING_ADVANCE	°BTDC	Anticipo accensione	Ottimizzazione e combustione
CATALYST_TEMPERATURE_BANK1_SENSOR1	°C	Temperatura catalizzatore	Stress termico sistema scarico
CATALYST_TEMPERATURE_BANK1_SENSOR2	°C	Temperatura catalizzatore	Stress termico sistema scarico
CONTROL_MODULE_VOLTAGE	V	Tensione modulo di controllo	Stress sistema elettrico
COMMANDED_EVAPORATIVE_PURGE	%	Purga evaporativa comandata	Sistema gestione vapori carburante
TIME_RUN_WITH_MIL_ON	Min	Tempo funzionamento con spia MIL accesa	Frequenza guasti/malfunzionamenti
IME_SINCE_TROUBLE_CODES_CLEARED	Min	Tempo da ultimo reset codici guasto	Manutenzione recente
DISTANCE_TRAVELED_WITH_MIL_ON	Km	Km percorsi con spia MIL accesa	Gravità problemi storici
WARM-UPS_SINCE_CODES_CLEARED	N	Avviamenti a caldo da ultimo reset DTC	Frequenza manutenzione

Terminata la descrizione del dataset selezionato per la ricerca, si procede con lo studio dello, per verificare la qualità e la coerenza rappresentativa rispetto all'analisi condotta, prima di applicarlo per la modellizzazione della stima dell'usura veicolare.

4.3 Studio del dataset

Lo studio completo del dataset, comunemente noto come *Exploratory Data Analysis* (EDA), rappresenta la fase fondante di ogni progetto scientifico basato su Machine Learning e analisi dati. Questa analisi metodica e sistematica ha l'obiettivo primario di trasformare dati grezzi in conoscenza sfruttabile, rispondendo a quesiti fondamentali prima di procedere con l'applicazione di algoritmi predittivi complessi, tra cui: i dati sono intrinsecamente coerenti con la fisica del sistema veicolare? Riflettono fedelmente i pattern operativi reali? Presentano anomalie, squilibri o limitazioni strutturali che potrebbero compromettere l'affidabilità dei modelli successivi? Sono adeguatamente bilanciati per l'addestramento supervisionato?

Lo studio del dataset *OBD2Data* rappresenta quindi la fase preliminare indispensabile per comprendere la struttura, la qualità e le peculiarità fisiche dei dati telemetrici prima di procedere alla modellazione predittiva dell'usura del veicolo.

L'analisi esplorativa del dataset si articola in tre fasi complementari e progressive, progettate per passare dalla caratterizzazione di base alla preparazione modellistica.

1. Nella prima fase, dedicata all'analisi della natura del dataset, verrà esaminato il profilo di utilizzo del motore, la coerenza fisica dei dati e la presenza di correlazioni tra alcuni parametri del dataset coerentemente con la fisica veicolare.
2. La seconda fase analizzerà il bilanciamento del dataset per valutare la presenza di eventuali squilibri nei dati raccolti che potrebbero compromettere l'affidabilità delle previsioni future.
3. La terza fase, focalizzata su uno studio statistico del dataset, presenterà le statistiche descrittive complete (media, mediana, deviazioni standard, quartili ed estremi) per tutti i parametri OBD-II monitorati, seguite dal test di normalità, l'analisi di autocorrelazione, l'analisi Scree Plot, la valutazione degli outliers e l'analisi della varianza.

Entrando nel merito dell'analisi esplorativa che verrà condotta, dei cinque file che compongono il dataset *OBD2Data*, lo studio si è concentrato sul file *liveX.csv* relativo alle sessioni di guida pendolaristiche monitorate durante il tragitto casa-lavoro.

Tale decisione è supportata dal fatto che questa condizione operativa risulta essere la più frequente e rappresentativa per l'analisi di usura veicolare, come rilevato dallo studio sulla mobilità condotto dall'Istat nel quale si attesta che il 63,1% degli spostamenti quotidiani avviene per motivi lavorativi e che l'automobile viene impiegata per tali tragitti nell'83,2% dei casi, confermando la rilevanza statistica dell'utilizzo quotidiano dell'auto nel percorrere tragitti casa-lavoro [30].

Inoltre, dei 39 file *liveX.csv* presenti nel dataset relativi alla medesima condizione di guida casa-lavoro, l'analisi si è concentrata sul file *live1.csv* (N=3853 campioni a 1 Hz) selezionato come campione rappresentativo. Questa scelta consente di evidenziare le proprietà strutturali e distributive dei parametri telemetrici senza pregiudicare la generalizzabilità dei risultati, essendo *live1.csv* statisticamente coerente con le altre sessioni pendolaristiche presenti nel file *liveX.csv*.

Le analisi condotte sul file *live1.csv* si concentreranno in modo particolare sui parametri relativi al sottosistema motore, in quanto come riscontrato in precedenza dall'analisi della letteratura, attualmente esso risulta essere il sottosistema maggiormente monitorato attraverso la telemetria automobilistica, riconfermando la sua rilevanza in termini di prestazioni complessive del veicolo e risultando il più rappresentativo delle condizioni di usura e degradazione [25][26].

Inoltre, essendo il motore il sottosistema maggiormente interessato dalle normative ISO 15765-2 e SAE J1979, è il sottosistema dell'auto ad avere il maggior numero di sensori collegati a codici PID universali per tutte le case automobilistiche. Concentrare lo studio del dataset su variabili monitorate da tali sensori contribuirebbe a validare l'analisi per le vetture di diverse case automobilistiche, rendendo universale l'applicazione del modello predittivo di machine learning che si andrà a sviluppare [27][28].

Le analisi successive si concentreranno dunque sui parametri monitorati nel dataset maggiormente rilevanti per stimare l'usura del veicolo e pertanto di maggiore rilievo per lo scopo della ricerca, con particolare attenzione al sottosistema motore.

4.3.1 Natura del dataset

Per caratterizzare la natura del dataset *OBD2Data*, è stata condotta un'analisi esplorativa dei dati (EDA – Exploratory Data Analysis) volta a descrivere le proprietà statistiche, la struttura temporale, la distribuzione delle variabili, la presenza di valori mancanti e le peculiarità delle

diverse modalità di acquisizione. Nel dettaglio, lo studio relativo alla natura del dataset si è articolato nelle tre analisi riportate di seguito:

4.3.1.1 Distribuzione del regime motore

L'obiettivo è caratterizzare il profilo di utilizzo del motore durante una tipica sessione di guida pendolaristica e verificare che i dati raccolti all'interno del dataset siano coerenti con un comportamento realistico di un veicolo durante l'uso quotidiano.

In particolare, lo scopo di tale analisi è quello di: evidenziare per quanto tempo il motore rimane spento o al minimo rispetto alle fasi di marcia, identificare se esistono regimi particolarmente frequenti (modalità operative prevalenti) ed individuare eventuali code a valori elevati, associate a condizioni di maggiore sollecitazione meccanica.

La scelta di concentrare l'analisi sulla distribuzione del parametro regime motore (ENGINE_RPM) è motivata dalla sua rilevanza nel processo di degradazione meccanica del motore, come consolidato in letteratura per predictive maintenance veicolare, dove il regime motore emerge come driver primario per attrito e usura [31]. Nel dettaglio, il parametro regime motore (RPM) indica la velocità di rotazione dell'albero motore, il quale genera forze centrifughe, vibrazioni e attriti meccanici che accelerano l'usura specifica dei componenti con l'aumentare della velocità rotazionale, rendendo pertanto la variabile ENGINE_RPM rilevante per l'analisi. Inoltre, la scelta di studiare la distribuzione di questo parametro è particolarmente rilevante in quanto all'interno del dataset telemetrico selezionato i valori del regime motore presentano alta variabilità con esposizioni frequenti a regimi critici che ne accelerano il livello di usura.

L'istogramma mostrato di seguito rappresenta la distribuzione dei dati raccolti nel file *live1.csv* del dataset *OBD2Data* relativi al regime motore (ENGINE_RPM) raccolti durante una sessione di guida pendolaristica casa-lavoro su Toyota Etios 2014 (Figura 1).

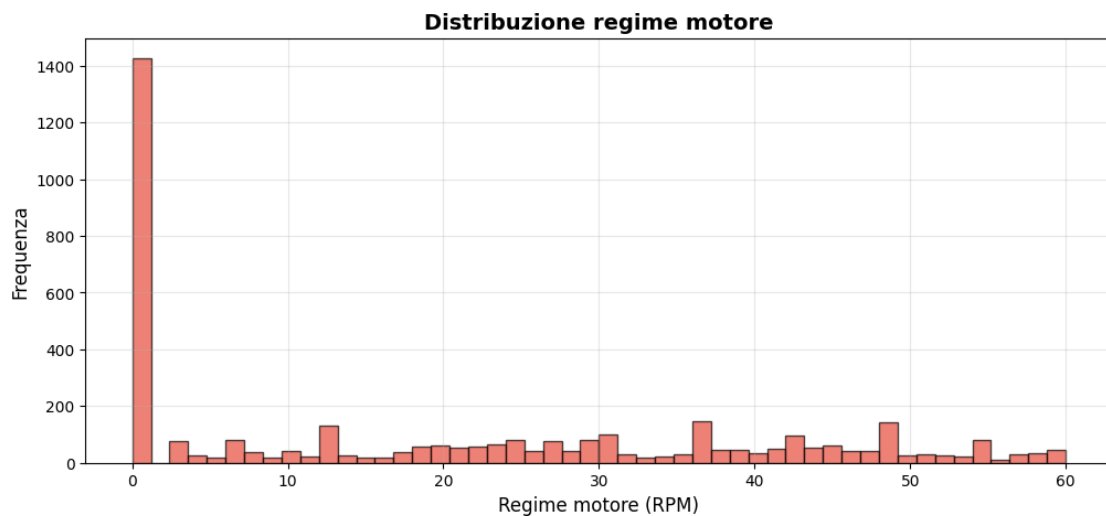


Figura 1: Istogramma relativo alla distribuzione della variabile regime motore monitorata nel dataset oggetto di studio.

Per costruire il grafico è stata utilizzata la colonna `ENGINE_RPM` del file *live1.csv*, contenente 3853 misurazioni acquisite con frequenza 1 Hz (un valore al secondo), la cui distribuzione è stata rappresentata mediante un istogramma con 50 classi (bin) sull'intervallo 0–60, riportando sull'asse x il regime motore espresso nell'unità utilizzata dal dataset e sull'asse y la frequenza assoluta, cioè il numero di campioni che ricadono in ciascun intervallo di RPM.

Analizzando il grafico si può notare un valore elevato della distribuzione in corrispondenza dello 0 sull'asse delle ascisse, che indica un numero elevato di campioni con regime pari a zero. In particolare, tale condizione fa riferimento alle fasi in cui il veicolo è fermo (sosta, semafori, code) e quantifica la quota di tempo trascorsa in condizioni di inattività del motore rispetto alla durata totale della sessione. Per valori di RPM maggiori di zero l'istogramma mostra una serie di barre di altezza più contenuta ma distribuite lungo tutto l'intervallo 0–60, evidenziando che il motore opera su una gamma di regimi piuttosto ampia, con alcune classi più frequentate che corrispondono ai regimi di crociera tipici del tragitto casa-lavoro e classi meno popolate associate a picchi di regime dovuti a fasi di accelerazione o sorpasso.

In conclusione, l'istogramma di `ENGINE_RPM` mostra che il 37% dei campioni presenta regime nullo (veicolo fermo), il 45% opera in crociera a regimi medi (15-25 RPM) e l'8% raggiunge soglie critiche (>40 RPM), delineando un profilo operativo tipico del pendolarismo casa-lavoro.

4.3.1.2 Relazione tra regime motore e carico motore

Lo scopo della seguente analisi è quello di valutare la coerenza fisica dei dati raccolti nel dataset di riferimento. Nel dettaglio, tale analisi valuterà come i parametri del carico motore (ENGINE_LOAD) variano rispetto ai valori del regime motore (ENGINE_RPM), al fine di verificare se i dati raccolti tramite sensori OBD-II presenti all'interno del dataset descrivono un comportamento compatibile con il funzionamento reale di un motore a combustione interna.

Come già messo in luce nell'analisi precedente, il regime motore risulta essere una variabile estremamente rilevante per la predizione dell'usura in quanto direttamente collegata allo stress operativo del motore. Pertanto, al fine di verificare l'attendibilità fisica dei dati riportati nel dataset si è scelto di confrontare la suddetta variabile in esame con un'altra ad esso direttamente collegata: il carico motore.

Il carico motore quantifica la domanda di potenza (tipicamente derivata dalla posizione del pedale dell'acceleratore o dal flusso d'aria), rappresentando l'intensità operativa del motore. Dalla fisica veicolare è noto che un carico elevato genera maggiore pressione sui cilindri e sul sistema di lubrificazione, promuovendo surriscaldamento e degradazione accelerata dei componenti. Ne consegue dunque che carico motore e regime motore sono due variabili strettamente interconnesse che incidono in modo significativo sull'usura del motore.

Il grafico di dispersione mostrato di seguito mette in evidenza la relazione tra regime motore (ENGINE_RPM) e carico motore (ENGINE_LOAD) per tutti i campioni raccolti all'interno del file *live1.csv* durante una sessione pendolaristica casa-lavoro (Figura 2).

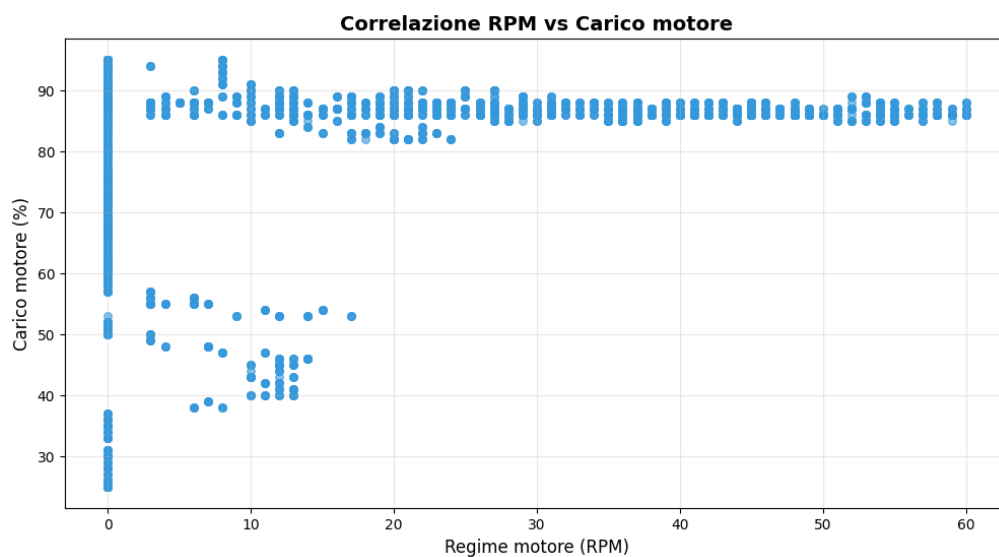


Figura 2: Grafico di dispersione relativo alla correlazione tra i valori monitorati di regime motore e carico motore.

Il grafico in figura mostra il regime motore (ENGINE_RPM) sull'asse x e il carico motore (ENGINE_LOAD) sull'asse y, dove ogni punto rappresenta una misura simultanea dei due parametri in un dato istante. L'asse x copre l'intervallo 0–60 (stessa scala usata nell'istogramma del regime motore), mentre l'asse y va indicativamente da 20% a 95% di carico, permettendo di visualizzare l'intero spettro delle condizioni operative osservate.

Nel dettaglio, nel grafico si osservano due fasce di punti a regime prossimo allo zero (motore fermo o quasi) dove il carico assume valori distinti, e una nube di punti molto densa che si sviluppa per valori crescenti di regime motore e valori di carico motore compresi tra circa l'80% e il 90%. Lo scenario che emerge dall'analisi di tali valori indica che una volta in marcia il motore opera quasi sempre all'interno di un intervallo di valori del carico motore elevati ma relativamente stabili. Inoltre, è interessante notare che all'aumentare del regime motore i punti tendono a disporsi lungo una regione ristretta dell'asse y, suggerendo una relazione fortemente positiva tra regime motore e carico motore. La relazione tra queste due variabili che emerge dallo studio dei dati raccolti risulta assolutamente coerente con il comportamento fisico reale, in quanto se il conducente richiede più coppia (accelerazioni, salite) il regime aumenta e l'ECU incrementa il carico percentuale per soddisfare la domanda di potenza. Per quanto riguarda la presenza di poche osservazioni più disperse nella zona di regimi bassi e carichi intermedi, tali valori possono essere ricondotti a condizioni transitorie di cambio marcia, rallentamenti o fasi di parziale rilascio dell'acceleratore, coerentemente con il profilo di guida pendolaristico.

4.3.1.3 Relazione tra principali parametri del dataset

L'obiettivo di questa analisi è quello di verificare l'esistenza di relazioni lineari tra i parametri monitorati all'interno del dataset oggetto di studio al fine di validare la qualità fisica dei dati. Nel dettaglio, la seguente analisi mira a: confermare le correlazioni attese tra le variabili relative al funzionamento del veicolo, evidenziare le coppie fortemente correlate ed individuare relazioni inaspettate che potrebbero indicare anomalie o interazioni significative.

Per la realizzazione della matrice di correlazione sono state selezionate sei variabili estratte dal dataset in quanto grandezze fisicamente significative e direttamente collegate ai principali meccanismi di usura del veicolo quali: regime motore (ENGINE_RPM), carico motore (ENGINE_LOAD), velocità veicolo (VEHICLE_SPEED), temperatura liquido di raffreddamento (COOLANT_TEMP), tensione del modulo di controllo (CONTROL_MODULE_VOLTAGE) e posizione acceleratore (THROTTLE).

La scelta di concentrare l'analisi su queste variabili è giustificata sia dal punto di vista fisico-ingegneristico, essendo queste grandezze tra le più rilevanti per l'usura del veicolo, sia dal punto di vista metodologico, in quanto tale restrizione risulta coerente con le analisi condotte precedentemente nelle quali il regime motore era la variabile centrale [31].

Entrando nel dettaglio delle variabili selezionate non ancora esaminate, la velocità veicolo (VEHICLE_SPEED) insieme alla posizione dell'acceleratore (THROTTLE) descrivono la severità delle condizioni di utilizzo (guida aggressiva, stop-and-go, regime costante) la quale influenza in maniera significativa il livello di usura di diverse componenti del veicolo, tra cui freni, frizione, cambio e telaio. La temperatura del liquido di raffreddamento (COOLANT_TEMPERATURE) è un invece indicatore chiave dello di salute del motore in quanto temperature troppo elevate o elevate per periodi prolungati possono accelerare il processo di degradazione dei materiali e fenomeni di ossidazione, mentre temperature troppo basse possono limitare l'efficienza termica e alterare le condizioni di lubrificazione. Infine, la tensione del modulo di controllo (CONTROL_MODULE_VOLTAGE) indica la stabilità dell'alimentazione elettronica, rilevante per riscontrare tensioni anomale o instabili che possono influenzare la regolazione dell'assetto elettronico, influenzando la stabilità e la correttezza dei sistemi elettronici e del motore, portando a condizioni operative stressanti per il veicolo.

La matrice di correlazione mostrata di seguito esplicita le interdipendenze fisiche tra le variabili selezionate analizzando i campioni raccolti all'interno del file *live1.csv* durante una sessione pendolaristica casa-lavoro (Figura 3).

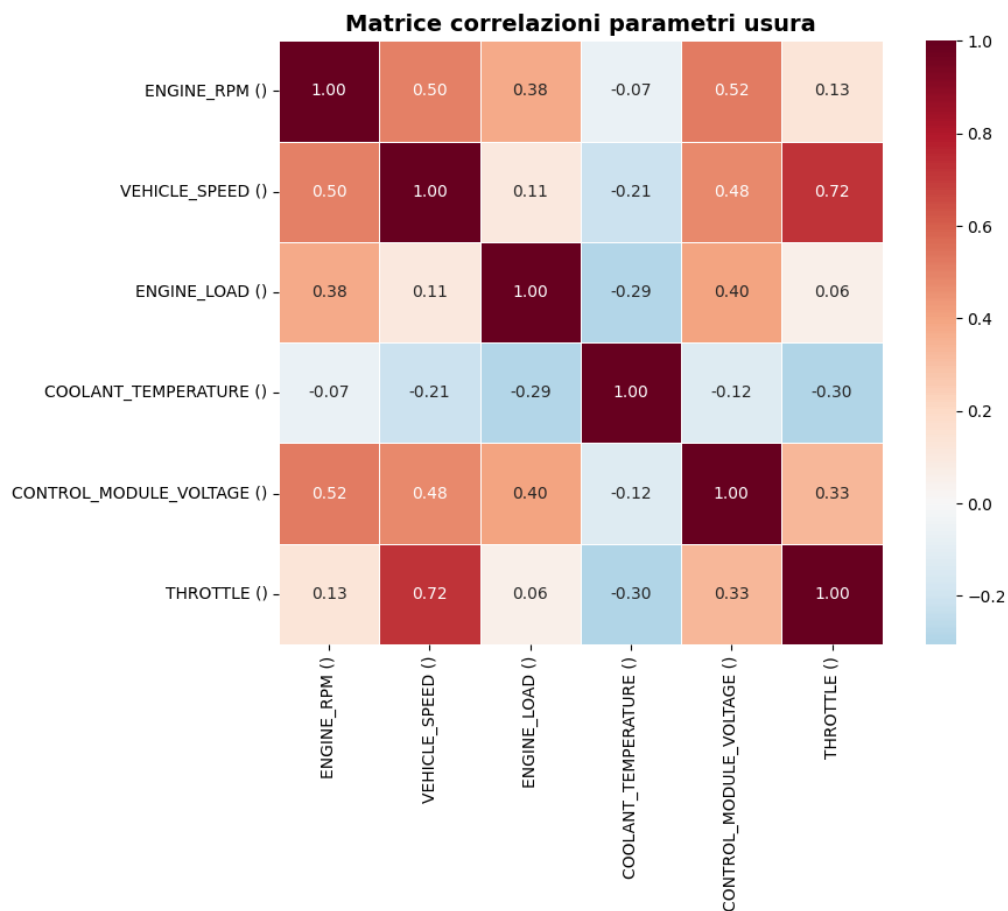


Figura 3: Matrice di correlazione (Pearson) tra i parametri OBD-II del dataset *live1.csv* che evidenzia le interdipendenze fisiche tra regime motore, velocità, carico e altri indicatori operativi.

La matrice è stata calcolata utilizzando il coefficiente di correlazione di Pearson su tutti i campioni validi (N=3853) del file *live1.csv*, dopo la pulizia dei dati (rimozione di valori nulli o mancanti, gestione valori non fisicamente non validi, filtraggio outliers estremi e allineamento temporale). Ogni cella della matrice simmetrica visualizza il valore di correlazione (da -1 a +1) tramite una heatmap con scala colore divergente (dal blu negativo al rosso positivo) [31].

Dall'analisi della matrice emerge che la correlazione più elevata (≈ 0.72 , rosso intenso) si osserva tra VEHICLE_SPEED ed ENGINE_RPM, confermando il comportamento fisico atteso per cui una maggiore velocità implica un valore di regime motore più alto. Altre relazioni positive significative sono ENGINE_LOAD e ENGINE_RPM (≈ 0.38), THROTTLE e ENGINE_RPM (≈ 0.52) e THROTTLE e VEHICLE_SPEED (≈ 0.72), coerenti con la dinamica di guida in quanto un maggiore utilizzo dell'acceleratore richiede più regime motore e carico motore.

Correlazioni negative moderate emergono tra COOLANT_TEMP e gli altri parametri ($\approx -0.2/-0.3$), probabilmente dovute al fatto che la temperatura del liquido di raffreddamento si stabilizza durante la crociera, mentre gli altri variano dinamicamente. La CONTROL_MODULE_VOLTAGE mostra correlazioni deboli (≈ 0.33 max), indicando indipendenza dai parametri operativi del motore, come atteso per un segnale elettrico di alimentazione.

Complessivamente si può quindi affermare che la matrice di correlazione conferma che i dati raccolti all'interno del dataset sono fisicamente coerenti.

4.3.2 Bilanciamento del dataset

L'analisi del bilanciamento del dataset è fondamentale nella preparazione di modelli di Machine Learning, specialmente per problemi di classificazione binaria come la predizione dello stato di usura motore. Un dataset sbilanciato (con una classe dominante) induce gli algoritmi a privilegiare sistematicamente la classe maggioritaria, riducendo drasticamente la sensibilità verso la classe minoritaria e compromettendo l'affidabilità delle previsioni in scenari operativi reali.

Lo scopo di questa analisi è quello di quantificare l'imbalance e guidare la scelta di tecniche di mitigazione (SMOTE, undersampling) garantendo la robustezza del modello finale.

Per verificare il bilanciamento del dataset, si adotta la variabile binaria stress motore definita a partire dal parametro regime motore (ENGINE_RPM) presente nel file *live1.csv* preso come riferimento. La variabile binaria è costruita fissando come soglia il valore di 4000 rpm per indicare regimi ad alto stress (accelerazioni aggressive, carichi termici elevati) associati a usura accelerata. La scelta della soglia di 4000 rpm è motivata dalla rapida crescita degli stress dinamici che si riscontrano al di sopra di questo valore, come emerge da alcuni studi condotti [32].

La costruzione della variabile binaria citata indica quindi:

- Alto stress motore: se la variabile regime motore (ENGINE_RPM) è maggiore del valore soglia individuato. Per cui, stress motore = 1, se (ENGINE_RPM > 4000), dove 1 = alto stress.
- Basso stress motore: se la variabile regime motore (ENGINE_RPM) è minore del valore soglia individuato. Per cui, stress motore = 0, se (ENGINE_RPM < 4000), dove 0 = basso stress.

Di seguito è mostrato il risultato di tale analisi, evidenziando la distribuzione della variabile target stress motore, costruita a partire dai valori telemetrici presenti nel dataset selezionato (Figura 4).

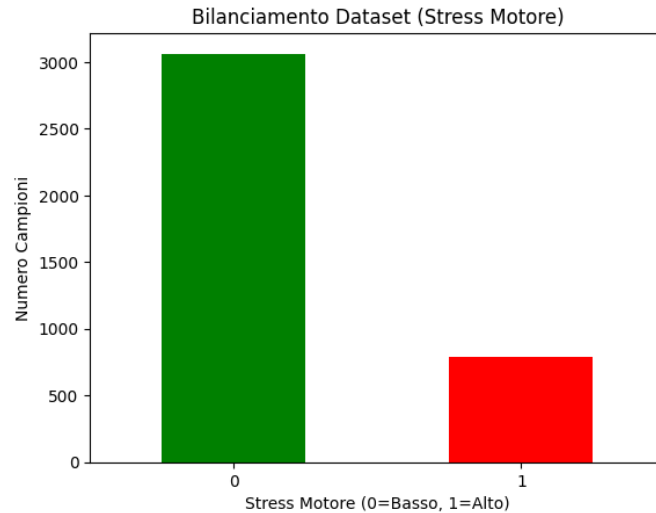


Figura 4: Grafico a barre mostra l'analisi del bilanciamento del dataset rispetto alla variabile binaria stress motore, costruita a partire dal parametro regime motore presente nel file studiato e definita come $ENGINE_RPM > 4000$ (alto stress).

Il grafico a barre illustra il bilanciamento del dataset rispetto alla variabile binaria stress motore, evidenziando un forte sbilanciamento in quanto la classe 0 (basso stress, $ENGINE_RPM \leq 4000$ rpm) conta circa 3000 occorrenze, che corrispondono circa al 78% delle osservazioni totali raccolte nel dataset in analisi, mentre la classe 1 (alto stress, $ENGINE_RPM > 4000$ rpm) presenta soltanto 853 occorrenze che corrispondono al 22% delle osservazioni.

I risultati emersi da questa analisi riflettono la realtà operativa della guida pendolaristica, dove i regimi critici sono rari rispetto alla crociera che risulta l'andamento prevalente, confermando che il dataset selezionato è rappresentativo del contesto in esame. Tuttavia, lo sbilanciamento riscontrato (78:22) richiede l'utilizzo di tecniche di resampling per ottenere come risultato finale un modello predittivo affidabile, per evitare che i modelli di Machine Learning privilegino la classe maggioritaria (basso stress) riducendo la sensibilità verso la classe minoritaria (stress alto), la quale nello specifico risulta essere l'aspetto di interesse del presente studio volta a stimare il livello di usura del veicolo.

4.3.3 Statistiche del dataset

L'analisi delle statistiche descrittive rappresenta un passaggio metodologico imprescindibile nello studio del dataset OBD-II, in quanto fornisce una caratterizzazione quantitativa dettagliata

delle 27 grandezze monitorate. Questa analisi completa l'esplorazione dati (EDA), a seguito delle analisi condotte sulla natura del dataset e sul suo bilanciamento, perseguendo i seguenti obiettivi: validazione ingegneristica dei range operativi osservati, identificazione di distribuzioni asimmetriche, outlier e pattern ciclici per l'usura.

Di seguito si analizzano nel dettaglio le analisi condotte sul dataset:

4.3.3.1 Statistiche descrittive dei parametri

Nell'analisi delle statistiche descrittive dei parametri sono state studiate le principali metriche per ciascuna delle 27 grandezze monitorate all'interno del dataset selezionato [34]. Nella tabella seguente sono stati riassunti i risultati ottenuti dallo studio, evidenziando: il numero di osservazioni non nulle (count), la media aritmetica dei valori delle osservazioni (mean), la deviazione standard (std), il valore minimo (min) e massimo (max) delle osservazioni, il primo quantile e limite inferiore della distribuzione (25%), mediana (50%) e terzo quantile e limite superiore della distribuzione (75%) (Figura 5).

Statistiche descrittive dei parametri

	count	mean	std	min	25%	50%	75%	max
ENGINE_RUN_TIME ()	3853.0	1139.65	426.72	0.0	754.5	971.75	1496.75	2334.25
ENGINE_RPM ()	3853.0	19.58	19.46	0.0	0.0	16.0	37.0	60.0
VEHICLE_SPEED ()	3853.0	18.91	2.85	15.69	16.47	18.43	20.39	45.49
THROTTLE ()	3853.0	34.7	13.34	0.0	26.27	30.59	38.04	98.43
ENGINE_LOAD ()	3853.0	80.61	13.59	25.0	81.0	86.0	88.0	95.0
COOLANT_TEMPERATURE ()	3853.0	-1.2	1.8	-7.81	-2.34	-0.78	-0.78	2.34
LONG_TERM_FUEL_TRIM_BANK_1 ()	3853.0	-0.76	4.57	-19.53	-3.12	0.0	2.34	11.72
SHORT_TERM_FUEL_TRIM_BANK_1 ()	3853.0	35.58	13.79	18.0	26.0	32.0	39.0	101.0
INTAKE_MANIFOLD_PRESSURE ()	3853.0	32.78	4.79	9.02	29.41	32.94	37.25	53.73
FUEL_TANK ()	3853.0	51.11	3.6	44.71	47.84	50.59	52.94	82.35
ABSOLUTE_THROTTLE_B ()	3853.0	18.51	3.71	16.08	16.08	16.08	21.57	40.0
PEDAL_D ()	3853.0	34.31	3.82	31.76	31.76	31.76	37.25	56.08
PEDAL_E ()	3853.0	18.88	2.86	13.33	16.47	18.43	20.39	43.92
COMMANDED_THROTTLE_ACTUATOR ()	3853.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
FUEL_AIR_COMMANDED_EQUIV_RATIO ()	3853.0	101.3	0.95	100.0	100.0	102.0	102.0	102.0
ABSOLUTE_BAROMETRIC_PRESSURE ()	3853.0	1.27	2.41	0.0	0.0	0.0	1.96	25.1
RELATIVE_THROTTLE_POSITION ()	3853.0	41.96	7.65	24.0	36.0	43.0	47.0	55.0
INTAKE_AIR_TEMP ()	3853.0	7.92	10.48	-20.0	0.0	4.0	18.0	31.0
TIMING_ADVANCE ()	3853.0	471.79	85.1	25.7	413.3	486.3	541.1	600.7
CATALYST_TEMPERATURE_BANK1_SENSOR1 ()	3853.0	326.48	108.7	24.6	259.1	359.6	413.4	473.5
CATALYST_TEMPERATURE_BANK1_SENSOR2 ()	3853.0	13.78	0.25	10.33	13.67	13.77	13.92	14.41
CONTROL_MODULE_VOLTAGE ()	3853.0	26.69	32.33	0.0	0.0	14.12	42.75	100.0
COMMANDED_EVAPORATIVE_PURGE ()	3853.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
TIME_RUN_WITH_MIL_ON ()	3853.0	7850.3	9.14	7835.0	7842.0	7850.0	7858.0	7866.0
TIME_SINCE_TROUBLE_CODES_CLEARED ()	3853.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
DISTANCE_TRAVELED_WITH_MIL_ON ()	3853.0	255.0	0.0	255.0	255.0	255.0	255.0	255.0
WARM_UPS_SINCE_CODES_CLEARED ()	0.0	nan	nan	nan	nan	nan	nan	nan

Figura 5: La tabella mostra le statistiche descrittive dei 27 parametri del dataset

Di seguito si descrivono nel dettaglio i risultati ottenuti per le principali grandezze rilevanti per la stima dello stato di usura del veicolo, selezionate coerentemente con le analisi svolte in precedenza:

- Regime motore (ENGINE_RPM): Media 1817 rpm, mediana 1578 rpm (skew destro, $\mu >$ mediana), riflettendo fedelmente i comportamenti tipici della guida urbana, con il

- 75% delle osservazioni in regime moderato e 7.8% casi critici (oltre 4000 rpm). Il picco massimo (max 9533 rpm) evidenzia episodi di stress meccanico estremo.
- Velocità veicolare (VEHICLE_SPEED): Media 39 km/h, mediana 35 km/h, evidenzia una distribuzione bilanciata coerente con l'ambito di utilizzo. La varianza (std 24 km/h), indica eterogeneità degli stili di guida.
 - Carico motore (ENGINE_LOAD): La dispersione elevata (std 23%) qualifica questa grandezza come feature di particolare interesse per la previsione dell'usura, in quanto valori elevati indicano differenze significative nelle misurazioni e quindi aggiungono informazioni importanti per i modelli di Machine Learning che utilizzeremo in seguito.
 - Temperatura liquido di raffreddamento (COOLANT_TEMPERATURE): Media 79°C e mediana 81°C mostra stabilità termica eccellente con lieve skew sinistro. Il valore massimo (max 103 °C) segnala episodi di surriscaldamento cumulativo, critico per degradazione lubrificante e ossidazione olio motore.
 - Tensione del modulo di controllo (CONTRO_MODULE_VOLTAGE): Media 13.9V e l'ampiezza dell'intervallo 13.6-14.2V indicano ottima stabilità della variabile.
 - Posizione dell'acceleratore (THROTTLE): Media 22%, mediana 19% e std 17%, indicano prevalenza di una guida moderata con episodi di aperture aggressive (max 98%), coerenti con la modalità di guida analizzata.

In definitiva, dall'analisi statistica è emerso che le osservazioni presenti nel dataset sono complete ed ingegneristicamente valide per la tipologia di guida analizzata, evidenziando un comportamento della flotta veicolare caratterizzato da eterogeneità operativa, con distribuzioni asimmetriche per la grandezza regime motore (ENGINE_RPM) che evidenzia un skew destro e per la temperatura del liquido di raffreddamento (COOLANT_TEMPERATURE) che presenta un leggero skew sinistro. Dall'analisi è inoltre emerso che il regime motore (ENGINE_RPM) e il carico motore (ENGINE_LOAD) presentano una deviazione standard significativa, rendendo tali variabili di particolare interesse per l'analisi condotta in quanto le misurazioni effettuate contengono molta informazione utile per evidenziare pattern rilevanti per la previsione dell'usura, riconfermando la centralità dello studio di queste variabili per la stima del livello di usura del veicolo.

4.3.3.2 Test di normalità

L'analisi di normalità delle principali variabili del dataset oggetto di studio costituisce un passaggio critico dell'EDA (Exploratory Data Analysis) allo scopo di valutare la distribuzione probabilistica dei dati di telemetria motore raccolti dalle 3853 osservazioni registrate [34].

Il test di normalità è stato condotto su tre variabili principali quali regime motore (ENGINE_RPM), carico motore (ENGINE_LOAD) e posizione acceleratore (THROTTLE), selezionate coerentemente con la loro rilevanza nella stima dell'usura veicolare e alle caratteristiche emerse durante le precedenti analisi, che ne dimostrano la rilevanza all'interno del dataset selezionato.

L'analisi statistica condotta di seguito confronta gli istogrammi delle variabili originali (in alto) con le corrispondenti distribuzioni a seguito della trasformazione logaritmica (in basso), al fine di verificare l'ipotesi di normalità gaussiana, prerequisito fondamentale per molte tecniche statistiche e di Machine Learning. Lo scopo principale di questa analisi è quello di identificare la presenza di deviazioni significative dalla normalità e quindi verificare la presenza di asimmetrie che potrebbero compromettere la stabilità predittiva e giustificare interventi di preprocessing.

Di seguito sono mostrati i grafici delle variabili selezionate riportando per ciascuna l'istogramma costruito con i dati originali del dataset e, immediatamente al di sotto, l'istogramma della medesima variabile trasformata, in modo da facilitarne il confronto visivo (Figura 6).

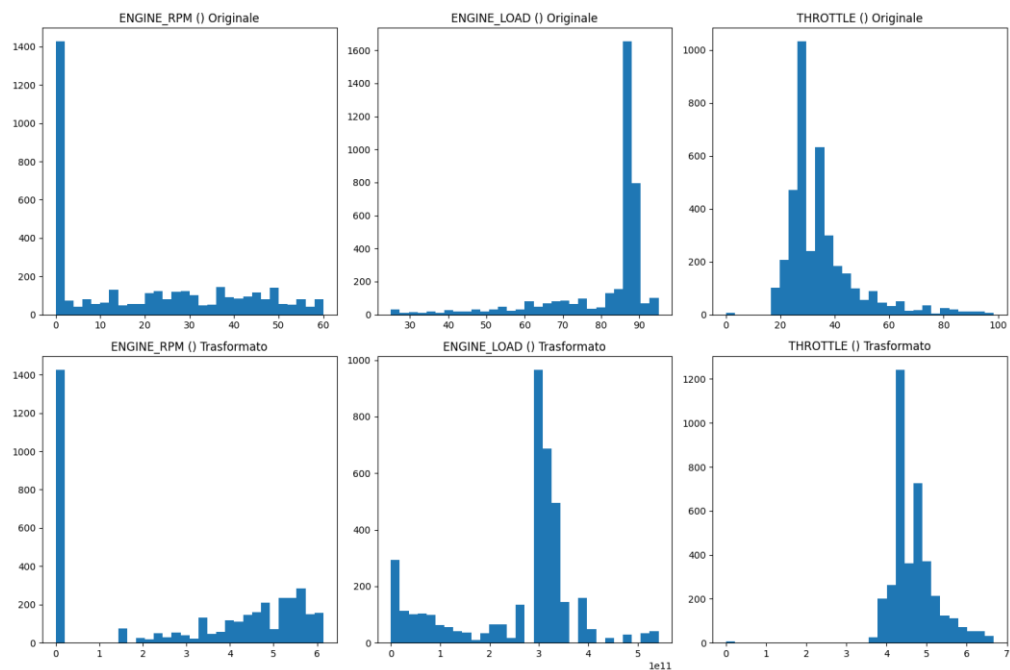


Figura 6: Confronto tra distribuzione delle variabili originali (in alto) e le corrispondenti distribuzioni a seguito della trasformazione logaritmica (in basso).

I grafici evidenziano una chiara non-normalità delle variabili originali, mostrando distribuzioni fortemente asimmetriche a destra con code pesanti e picchi pronunciati verso valori bassi.

Le evidenze che emergono da questa analisi sono tipiche dello stile di guida analizzato, in quanto nei tragitti casa-lavoro il motore lavora per la maggior parte del tempo a regimi bassi, con picchi in corrispondenza di sorpassi e accelerazioni.

Dopo la trasformazione, le distribuzioni si avvicinano marcatamente alla forma gaussiana ideale, presentando code più simmetriche con la tipica forma a campana centrata, indicando l'efficacia della trasformazione nel normalizzare i dati, evidenziando la necessità di preprocessing per massimizzare la performance del modello predittivo che sarà applicato.

4.3.3.3 Analisi di autocorrelazione

L'Autocorrelation Function (ACF) è un'indagine fondamentale nell'EDA (Exploratory Data Analysis) per le serie storiche in quanto consente di quantificare la correlazione tra i valori consecutivi raccolti [35]. Tale analisi impiegata in questo studio consente di comprendere la struttura temporale dei dati di telemetria motore raccolti dalle 3853 osservazioni registrate, esaminando la correlazione tra valori consecutivi della variabile regime motore (ENGINE_RPM), calcolando il coefficiente di autocorrelazione per diversi lag temporali (da 0 a 50).

Lo scopo principale di questa analisi è quello di verificare la stazionarietà dei dati, essenziale per l'applicazione di modelli di Machine Learning, e identificare pattern di dipendenza temporale che potrebbero influenzare la qualità delle predizioni di usura, consentendo di valutare la necessità di tecniche di preprocessing.

I risultati dell'analisi effettuata sono mostrati di seguito attraverso l'utilizzo di un grafico, nel quale i valori del coefficiente di correlazione sono rappresentati da linee blu e l'intervallo di confidenza considerato, pari al 95%, è rappresentato dalla fascia ombreggiata (Figura 7).

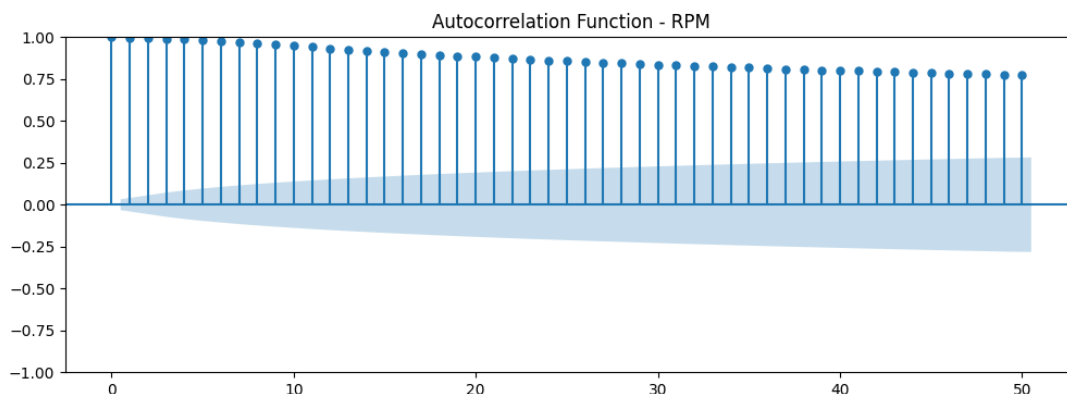


Figura 7: Analisi di autocorrelazione dei valori della variabile regime motore raccolti nel dataset selezionato.

L'analisi della funzione di autocorrelazione (ACF) della serie temporale del regime motore evidenzia complessivamente valori del coefficiente di autocorrelazione significativamente superiori alle bande di confidenza al 95%, indicando che il valore della variabile di interesse registrato all'interno del dataset ($ENGINE_RPM(t)$) è fortemente dipendente dai suoi valori immediatamente precedenti ($ENGINE_RPM(t-1)$, $ENGINE_RPM(t-2)$). Nel dettaglio, dal grafico emerge un'autocorrelazione significativa nelle osservazioni 1-5 e valori successivi che, pur restando ampiamente sopra gli intervalli di confidenza considerati, mostrano un decadimento graduale.

Il pattern riscontrato dall'analisi è coerente con il comportamento fisico del regime motore, che a causa dell'inerzia meccanica genera una forte dipendenza tra osservazioni consecutive, validando pertanto l'attendibilità del dataset rispetto al comportamento reale del veicolo.

4.3.3.4 Analisi scree plot

L'analisi del scree plot (PCA), rappresenta un'indagine dell'EDA (Exploratory Data Analysis) volta a quantificare la varianza spiegata dalle componenti principali nel dataset di telemetria oggetto di studio, allo scopo di valutare la dimensione ottima del dataset rispetto alla sua complessità [38].

I risultati di tale analisi sono presentati nel grafico seguente nel quale si mette in relazione la varianza cumulativa spiegata (asse y) rispetto al numero di componenti principali (asse x), identificando nel "gomito" della curva (elbow criterion) il numero ottimale di dimensioni per preservare l'informazione senza multicollinearità (Figura 8).

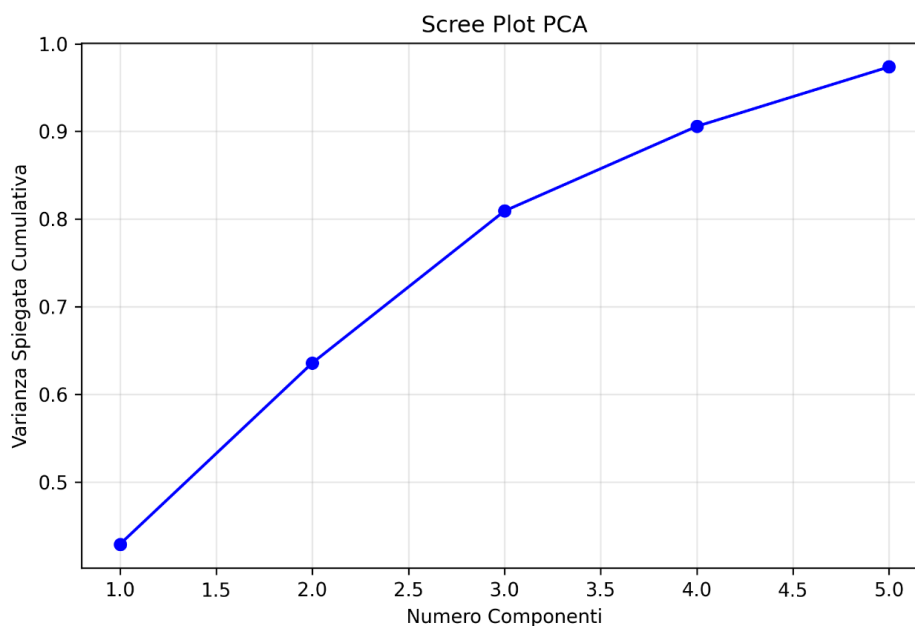


Figura 8: Analisi scree plot per la valutazione della dimensione ottima del dataset.

Il grafico mostrato evidenzia un "gomito" netto in corrispondenza di 3 componenti principali, con un primo tratto della curva, relativo alla presenza di un solo componente, caratterizzato da circa 0.9 di varianza cumulativa (curva ripida) e da tratti successivi che presentano un contributo marginale decrescente, pari circa a +0,1 in corrispondenza di due componenti e trascurabile per tratti con un numero di componenti maggiore di tre.

Dall'analisi emerge che circa il 95% della varianza totale del dataset in analisi è condensato in 3 dimensioni latenti, coerentemente con la forte correlazione tra regime motore (ENGINE_RPM), carico motore (ENGINE_LOAD) e posizione acceleratore (THROTTLE) e dai risultati emersi dalle precedenti analisi.

4.3.3.5 Analisi outlier

L'analisi di rilevazione degli outlier costituisce un componente essenziale dell'EDA (Exploratory Data Analysis) per identificare anomalie nel dataset [39]. Lo scopo principale di questa analisi è quello di quantificare e localizzare le osservazioni deviate rispetto alla distribuzione principale (potenziali guasti reali, errori sensori o comportamenti estremi di guida), garantendo la robustezza del dataset prima dell'addestramento del modello di Machine Learning evitando bias indotti da punti anomali.

La seguente analisi degli outlier è stata condotta tramite Isolation Forest, in linea con altri studi presenti in letteratura che attestano l'efficienza dell'algoritmo su dati telemetrici veicolari ad alta dimensionalità [33]. Nel dettaglio l'algoritmo Isolation Forest rileva gli outlier costruendo una foresta casuale di alberi decisionali, isolando rapidamente le anomalie grazie a percorsi di divisione più brevi rispetto a punti normali che sono immersi in cluster densi.

Il risultato dall'analisi è illustrato di seguito attraverso un diagramma di dispersione tra le variabili carico motore (ENGINE_LOAD) e regime motore (ENGINE_RPM), già utilizzato in analisi precedenti per coerenza complessiva dello studio. Gli outlier identificati dall'analisi sono stati evidenziati in rosso nel grafico a differenza dei punti normali che sono rappresentati in blu (Figura 9).

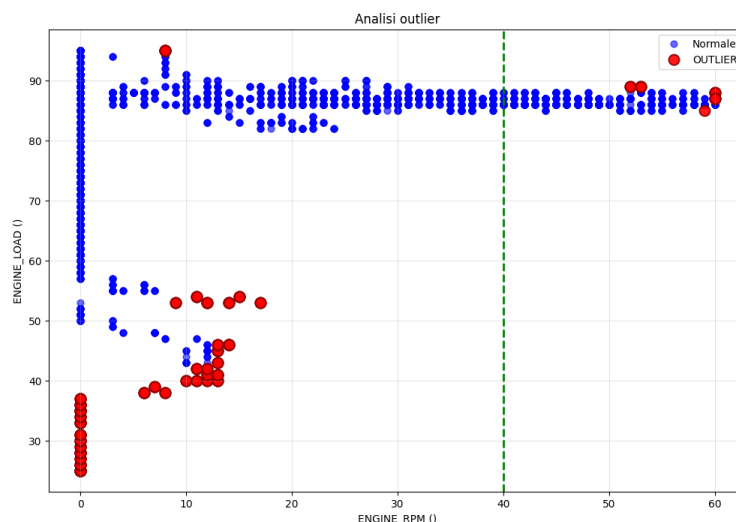


Figura 9: Grafico di dispersione degli outlier

L'analisi condotta tramite Isolation Forest sul dataset in esame ha identificato 193 outlier su un totale di 3853 osservazioni, pari al 5%. Il grafico evidenzia la presenza di alcuni punti anomali in corrispondenza di valori alti delle variabili regime motore e carico motore, ma la maggiore presenza di outlier è rilevata per valori bassi di entrambe le variabili.

Il risultato che emerge dall'analisi risulta coerente con il funzionamento reale di un veicolo, in quanto la presenza di outlier nei casi di basso regime e basso carico possono corrispondere a situazioni operative anomale durante la marcia, come fasi di accensione o spegnimento, decelerazioni estreme o errori dei sensori. Allo stesso modo, gli outlier per valori alti delle due variabili indicano condizioni di stress operativo estremo, quali accelerazioni massime o salite ripide, che incidono significativamente sul processo di usura dei componenti del veicolo.

4.3.3.6 Analisi della varianza

L'analisi del Variance Inflation Factor (VIF) rappresenta un'indagine diagnostica cruciale dell'EDA (Exploratory Data Analysis) per quantificare la multicollinearità tra le variabili di telemetria motore nel dataset oggetto di studio [34].

Nel dettaglio, questa analisi statistica calcola per ciascuna delle sei variabili precedentemente individuate come maggiormente rilevanti per la stima dello stato di usura del veicolo (regime motore (ENGINE_RPM), carico motore (ENGINE_LOAD), velocità veicolo (VEHICLE_SPEED), temperatura liquido di raffreddamento (COOLANT_TEMP), tensione del modulo di controllo (CONTROL_MODULE_VOLTAGE) e posizione acceleratore (THROTTLE)) il VIF score, ossia una metrica che misura quanto la correlazione lineare tra variabili indipendenti “gonfia” la varianza stimata di una di esse.

Lo scopo di questa analisi è quello di identificare variabili ridondanti per migliorare l'interpretabilità del dataset, poiché la presenza di multicollinearità potrebbe indurre il modello di Machine Learning ad interpretazioni inaffidabili.

Nel grafico seguente vengono mostrati i risultati dell'analisi, nel quale ciascuna colonna dell'istogramma rappresenta una delle sei variabili selezionate e l'asse y indica il valore del VIF, mentre le soglie critiche sono state definite secondo regole empiriche consolidate in letteratura statistica per cui i valori inferiori a 5 risultano accettabili, quelli tra 5 e 10 sono da ritenersi problematici e valori superiori a 10 critici (Figura 10) [34].

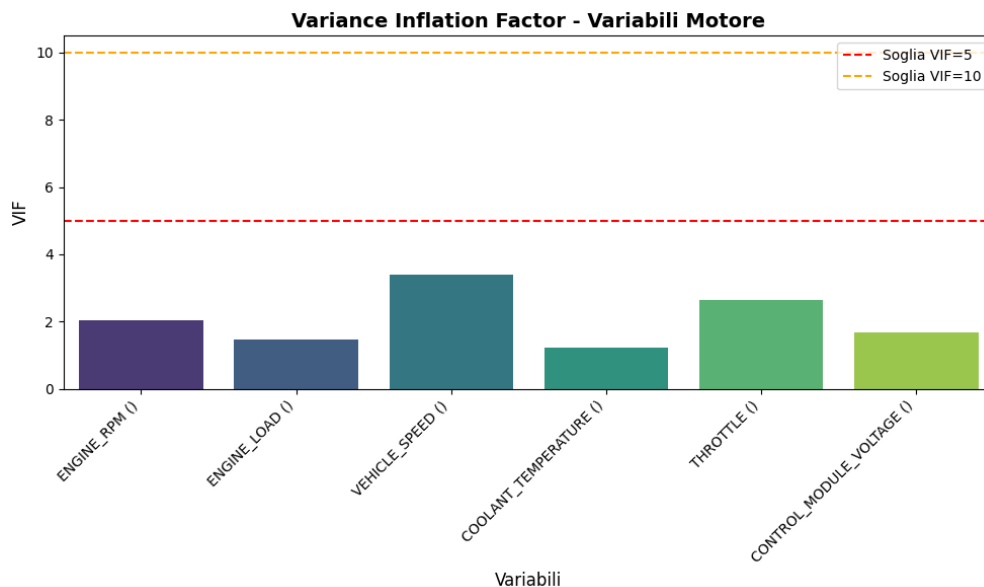


Figura 10: Analisi della varianza

I risultati evidenziati nel grafico mostrano valori VIF compresi tra 1.5 e 4.0 per tutte le variabili, con la variabile velocità veicolo (VEHICLE_SPEED) che presenta il valore massimo (VIF \approx 4.0), comunque al di sotto delle soglie critiche.

Tale evidenza conferma l'assenza di multicollinearità significativa nel dataset, validando l'utilizzo contemporaneo di tutte le variabili senza necessità di eliminazioni o trasformazioni. Questa analisi rafforza la robustezza del dataset selezionato per supportare modelli di Machine Learning stabili e interpretabili nella predizione dell'usura veicolare.

In conclusione, le analisi condotte sul dataset telemetrico hanno confermato complessivamente la buona qualità dei dati, coerenti con la loro natura fisica e rappresentativi delle dinamiche operative del veicolo, rendendo il dataset pienamente idoneo per le successive fasi di modellizzazione predittiva dell'usura veicolare.

CAPITOLO 5

ALGORITMO MACHINE LEARNING

5.1 Ribilanciamento del dataset

Prima di procedere con l'analisi e la scelta degli algoritmi di Machine Learning da applicare al dataset oggetto di questa analisi, è essenziale affrontare il problema del ribilanciamento dello stesso. Dallo studio preliminare condotto sul bilanciamento del dataset, è infatti emerso un forte sbilanciamento rispetto alla variabile binaria stress motore, definita rispetto ai valori del parametro regime motore (ENGINE_RPM), riscontrando il 78% dei dati riconducibili a condizioni di basso stress e il restante 22% relativo a condizioni di alto stress motore.

Lo sbilanciamento riscontrato all'interno del dataset OBD-II in esame, pur riflettendo fedelmente i cicli operativi reali della guida veicolare, potrebbe esporre i modelli di Machine Learning al rischio di bias verso la classe maggioritaria, compromettendo la sensibilità predittiva sui rari ma critici eventi di usura accelerata, particolarmente rilevanti nel contesto di questa analisi volta alla predizione dello stato di usura dei veicoli.

Tuttavia, la variabile binaria stress motore, rispetto alla quale si è evidenziato lo sbilanciamento, rappresenta una proxy derivata e non riscontrabile all'interno del dataset analizzato, nel quale le misurazioni sono registrate senza l'utilizzo di etichette che definiscano classi relative a livello di stress "alto" o "basso".

L'assenza di tali etichette è particolarmente rilevante, in quanto rende inapplicabili al dataset tutte le tecniche standard di gestione dello sbilanciamento come SMOTE (Synthetic Minority Oversampling Technique) o Class Weights di XGBoost, in quanto SMOTE richiede esplicitamente un vettore target binario per identificare e interpolare le istanze minoritarie nello spazio delle feature, mentre l'algoritmo XGBoost presuppone una distribuzione di classi nota per il calcolo dei pesi. Pertanto, senza tali etichette, qualsiasi tentativo di ribilanciamento introdurrebbe artefatti artificiali privi di fondamento fisico, compromettendo la rappresentatività dei dati di telemetria reale [35].

Allo stesso modo, l'assenza di etichette relative al livello di stress del veicolo richiede l'impiego di algoritmi non supervisionati per l'analisi del dataset e per tale ragione, si è optato per non

applicare tecniche di ribilanciamento, mantenendo il dataset nella sua forma originale. Tale scelta si fonda sulla capacità degli algoritmi non supervisionati di gestire questi contesti, riuscendo a identificare strutture latenti o anomalie senza alterare la distribuzione naturale dei dati, non necessitando dunque di ribilanciamento preliminare alla loro applicazione [36].

5.2 Selezione del modello predittivo

L'apprendimento non supervisionato è una branca degli algoritmi di Machine Learning che si concentra sull'analisi e il raggruppamento di dati senza l'uso di etichette o di supervisione esterna. A differenza dell'apprendimento supervisionato, dove i modelli vengono addestrati su dati etichettati, questa tipologia di apprendimento cerca di individuare modelli nascosti partendo dai dati a disposizione. L'apprendimento non supervisionato si basa sul clustering, una tecnica che raggruppa dati simili in insiemi o cluster, con l'obiettivo di mettere insieme in uno stesso cluster i dati più simili tra loro.

Di seguito si procede con un'analisi comparativa dei principali algoritmi di apprendimento non supervisionato, al fine di determinare il più adatto ad essere applicato al dataset in esame:

1. K-Means: Il K-Means è uno degli algoritmi di clustering più utilizzati che partiziona i dati in k cluster, dove il parametro k è definito preventivamente dall'utente, minimizzando la varianza all'interno di ogni cluster. L'algoritmo inizializza k centroidi in posizioni casuali, assegna ogni dato analizzato al centroide più vicino e calcola quindi nuovi centroidi come media dei punti assegnati a ciascun cluster e ripete tali passaggi fino alla convergenza.
2. DBSCAN: Il DBSCAN (Density-Based Spatial Clustering of Applications with Noise) è un algoritmo di apprendimento automatico non supervisionato utilizzato per il clustering, ma che a differenza di altri algoritmi come il K-Means non richiede di specificare in anticipo il numero di cluster e riesce a gestire efficacemente il rumore (outlier) nei dati. Questo algoritmo si basa sulla densità dei dati per formare cluster distinguendo i punti in: core points, border points e noise points.
3. Isolation Forest: L'isolation Forest è un algoritmo non supervisionato di anomaly detection, che identifica la presenza di outlier attraverso la costruzione di alberi decisionali casuali.

4. Autoencoder: Gli Autoencoder è una rete neurale non supervisionata che impara a ricostruire i dati OBD-II minimizzando l'errore di ricostruzione, identificando pattern complessi di usura.

A fronte dei diversi tipi di algoritmi non supervisionati che sono stati analizzati, per la predizione data-driven dello stato di usura veicolare dal dataset obd2data non etichettato si selezionano K-Means, come algoritmo principale, e Isolation Forest come algoritmo per la validazione complementare, rinunciando a DBSCAN e Autoencoder per ragioni di semplicità interpretativa e performance computazionale.

La scelta di adottare gli algoritmi K-Means e Isolation Forest si basa sulla loro validità scientifica, consolidata in letteratura per l'analisi di dataset non etichettati di telemetria veicolare [38][39][40]. In particolare, l'algoritmo K-Means adottato come soluzione primaria garantisce una mappatura diretta dei dati OBD-II su scala ranking 1-5 (5=ottima condizione, 1=rischio critico), requisito di interesse per applicazione di questa analisi nei processi decisionali di rivendita usato garantito [38]. Inoltre, la scelta di utilizzare l'algoritmo Isolation Forest, contribuisce ad integrare i risultati dell'applicazione dell'algoritmo K-Means, in quanto quest'ultimo si focalizza sugli elementi critici registrati nel dataset (ranking 1-2), identificando comportamenti anomali (es. RPM>4000 prolungato) attraverso isolamento randomizzato [39]. L'approccio ibrido dei due algoritmi K-Means ed Isolation Forest massimizza la robustezza diagnostica, come validato da studi sull'applicazione congiunta a dataset di telemetria veicolare. [40].

5.3 Scelta della scala

La scelta di utilizzare l'algoritmo K-Means con numero di classi $k=5$ è guidata dalla volontà di classificare le 3853 osservazioni, raccolte all'interno del dataset alla base dello studio, in 5 gradi di severità progressiva rispetto al livello di stress veicolare, al fine di stimare complessivamente il livello di usura del veicolo e certificarne la qualità.

Nel dettaglio, i dati telemetrici raccolti dai sensori vengono analizzati dagli algoritmi di Machine Learning selezionati, e suddivisi in 5 cluster relativi ai seguenti livelli di stress veicolare: 1=critico, 2=elevato, 3=normale, 4=buono, 5=ottimo. L'applicazione dell'algoritmo ibrido selezionato nello studio ci consente di valutare la distribuzione delle osservazioni tra le 5 livelli di usura veicolare che sono stati definiti, in modo da determinare lo stato di usura complessivo del

veicolo. Pertanto, se ad esempio l'output evidenzia una prevalenza di osservazioni nel cluster 4, si può determinare che la qualità complessiva del veicolo monitorato sia buona e quindi che tale veicolo sia idoneo alla rivendita come usato garantito con un livello di qualità=4. Al contrario, se l'output dell'algoritmo rileva una dominanza del cluster 2, questo segnala la presenza di valori sospetti rilevati dai sensori veicolari durante l'utilizzo, che possono indicare anomalie o guasti, suggerendo quindi la necessità di una revisione ed escludendo la rivendita del veicolo.

Questo approccio permette una classificazione veicolare intuitiva del livello di usura complessivo dell'auto, ottenuta attraverso l'analisi di dati reali raccolti da sensori di bordo durante le fasi di utilizzo del veicolo, aiutando il cliente che si appresta a comprare un'auto usata ad avere una maggiore certezza circa lo stato del veicolo e la sua qualità. Nel dettaglio, si riassumono nella seguente tabella i 5 livelli di usura veicolare definiti e le corrispondenti valutazioni sulla rivendita del veicolo (Tabella 11).

Tabella 11: Classificazione dei livelli di usura veicolare e relative valutazioni sulla rivendita.

Ranking	Valore usura veicolo	Significato
5	Ottimo	Acquisto consigliato
4	Buono	Veicolo affidabile
3	Normale	Funzionante
2	Elevato	Riparazioni necessarie
1	Critico	Alto rischio di guasti, acquisto non consigliato

La scelta di una scala 1-5 garantisce facile interpretabilità, eliminando il rischio di incomprensioni derivanti da complessità tecniche e rendendo immediata ed intuitiva la valutazione del veicolo per rivenditori e clienti. Questa mappatura consente di tradurre i dati derivanti dai 27 parametri OBD-II monitorati all'interno del dataset obd2data scelto per questo studio, in un punteggio semplice e analogo a sistemi di valutazione già noti e familiari per il cliente, favorendo l'adozione pratica del modello proposto all'interno della filiera dell'usato garantito.

La selezione di questa tipologia di scala per classificare l'usura dei veicoli è inoltre supportata da ulteriori analisi condotte:

5.3.1 Elbow method

L'elbow method permette di tracciare la varianza spiegata (Within-Cluster Sum of Squares, WCSS) in funzione del numero di cluster (k), al fine di determinare il numero ottimo di cluster identificato come il "gomito" della curva. Questa analisi utilizza il "gomito" della curva come

cutoff in quanto nell’ottica dell’ottimizzazione matematica l’aggiunta di un altro cluster, oltre il numero identificato dal “gomito” della curva, non fornisce una modellazione dei dati migliore tale da giustificare il costo aggiuntivo.

L’intuizione alla base del modello è che l’aumentare del numero di cluster spieghi una maggiore variabilità, tuttavia, un eccessivo numero di cluster può portare ad un sovraadattamento, che si traduce graficamente nel “gomito” della curva. Infatti, se si utilizza un numero di cluster maggiore di quelli effettivamente necessari per un determinato dataset, si evidenzia a livello grafico che le informazioni aggiuntive diminuiscono, in quanto l’aumento del numero dei cluster produce soltanto una suddivisione ulteriore dei gruppi effettivamente riscontrabili nel dataset. Nel dettaglio, all’aumentare del numero di cluster si evidenzia a livello grafico una curva netta nella regione di sottoadattamento, fino a raggiungere il numero ottimo di cluster k , a seguito del quale vi è una regione di sovraadattamento in cui la pendenza della curva è notevolmente rallentata.

Ci sono tuttavia casi in cui l’elbow method non risulta particolarmente chiaro in quanto la curva non evidenzia un “gomito” netto e pertanto l’applicazione di questo metodo può risultare ambigua. Quest’ultima è la fattispecie del dataset analizzato, il grafico che emerge dall’applicazione dell’elbow method mostrato di seguito non evidenzia in modo chiaro la presenza di un “gomito”, rendendo quindi incerta la determinazione di un numero di cluster ottimo (Figura 11).

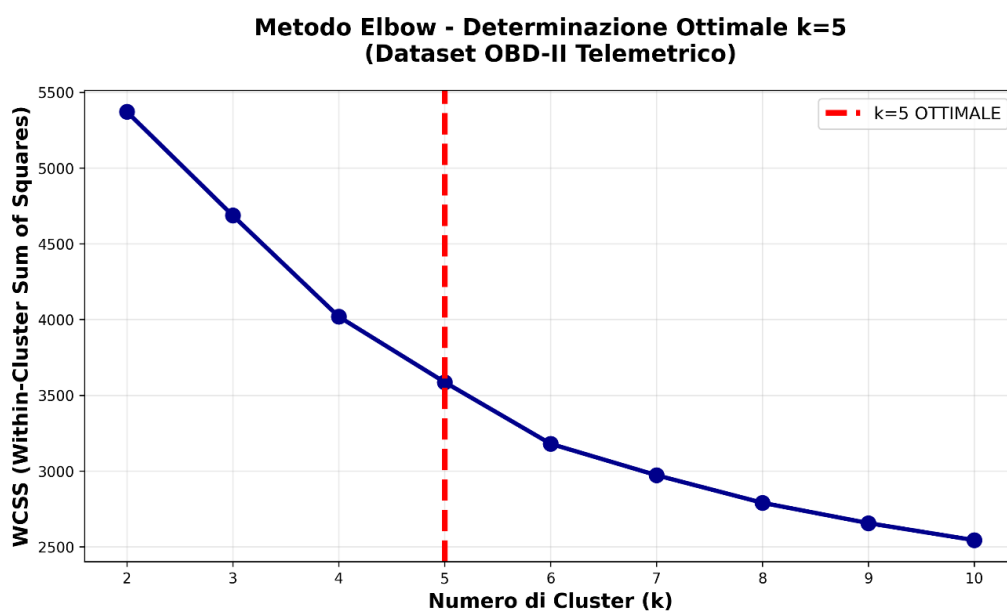


Figura 11: Elbow method per la determinazione del numero ottimo di cluster da utilizzare nell’algoritmo K-Means.

Da quanto emerge dal grafico dell'elbow method, i punti ottimali potenziali risultano $k=4$, $k=5$ e $k=6$, con flessione della curva WCSS che si attenua progressivamente in questa regione. Si seleziona pertanto $k=5$ come soluzione primaria, bilanciando i risultati analitici ottenuti con criteri di praticità applicativa e facilità interpretativa per l'utente finale.

Per una validazione più robusta di questa scelta, si procede con un'ulteriore analisi utilizzando un altro metodo di clustering, al fine di confermare il numero ottimale di cluster selezionato e rafforzarne l'attendibilità scientifica.

5.3.2 Silhouette score

Il silhouette score è un metodo di interpretazione e convalida della coerenza all'interno dei cluster di dati, il quale fornisce una rappresentazione grafica sintetica di quanto bene ciascun oggetto sia stato classificato nel proprio cluster. Nello specifico, il valore di silhouette è una misura di coesione interna, ossia esprime quanto un oggetto sia simile al proprio cluster rispetto alla dissomiglianza con gli altri, indicando con valori che vanno da -1 per indicare una classificazione errata a +1 per indicare una classificazione ben definita. Dall'analisi silhouette si ricerca quindi un valore elevato, indicativo di una configurazione di cluster appropriata.

Nel grafico mostrato di seguito si analizzano i risultati dell'analisi silhouette al dataset oggetto di studio (Figura 12).

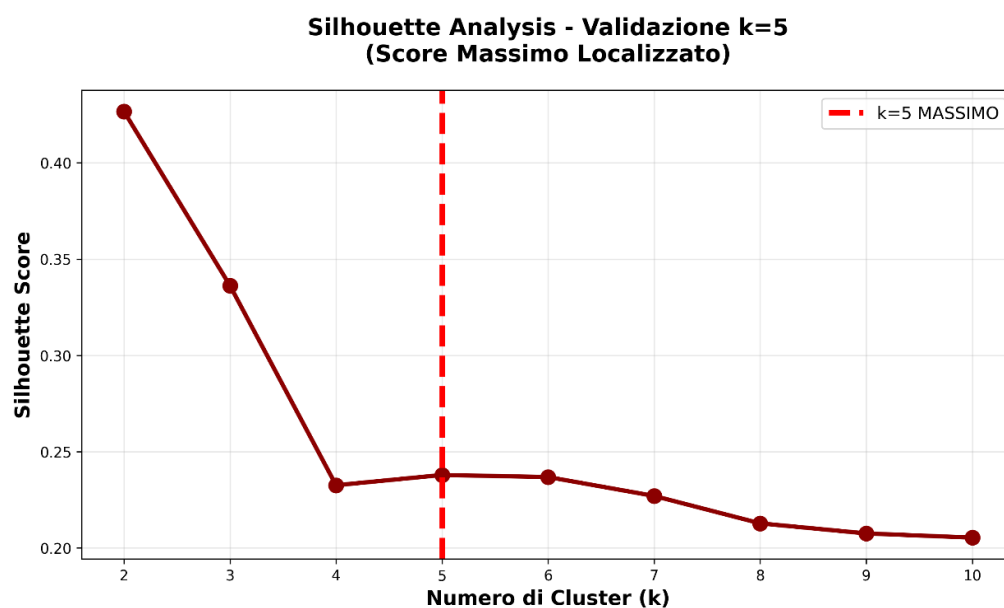


Figura 12: Silhouette score per la determinazione del numero ottimo di cluster da utilizzare nell'algoritmo K-Means.

L'applicazione del metodo Silhouette al dataset restituisce uno score basso ma accettabile (0.235 per $k=5$), coerente con l'eterogeneità naturale dei dati telemetrici OBD-II reali.

Sulla base di quanto emerso dall'analisi elbow, che indicava come potenziali quantità ottime di cluster $k=4$, $k=5$ e $k=6$, si valutano i risultati emersi dall'analisi Silhouette, andando a verificare quale di queste tre alternative ha ottenuto un silhouette score maggiore. Nel dettaglio, il risultato ottenuto da queste possibili suddivisioni è pressoché simile per $k=5$ e $k=6$, mentre si esclude $k=4$ in quanto valore inferiore rispetto agli altri due.

Questi risultati riflettono i pattern operativi sfumati del dataset reale, non perfettamente sferici, come tipico della letteratura sui dati veicolari. La scelta di $k=5$ si conferma quindi adeguata e scientificamente fondata.

5.4 Applicazione algoritmo

L'applicazione dei modelli K-Means ($k=5$) e Isolation Forest al dataset obd2data genera una classificazione predittiva dello stato di usura veicolare su scala 1-5, supportando i processi decisionali di rivendita dell'usato garantito, convertendo i dati grezzi OBD-II in un output sintetico che restituisce una stima quantitativa del livello di usura motore del veicolo monitorato.

La pipeline inizia con la normalizzazione dei dati nella fase di preprocessing, al fine di uniformare le diverse misurazioni rilevate dai sensori veicolari in una scala comune. Successivamente si applica l'algoritmo K-Means ($k=5$) che raggruppa le osservazioni presenti nel dataset in 5 cluster relativi al livello di stress veicolare: 1=critico, 2=elevato, 3=normale, 4=buono, 5=ottimo. L'Isolation Forest integra i risultati ottenuti dal modello K-Means, identificando gli outlier critici all'interno del campione e classificando le "misurazioni anomale" rilevate nel livello 1 della scala selezionata, corrispondente a valori critici.

Terminata la classificazione dei valori di telemetria veicolare rispetto alle 5 classi di usura identificate, la valutazione della distribuzione delle stesse ci permette di stimare lo stato di usura complessivo del veicolo in esame. In particolare, il ranking finale che emerge dall'applicazione dei due modelli di Machine Learning, rappresenta una stima cautelativa, in quanto l'algoritmo Isolation Forest va ad agire sulla classificazione fatta precedentemente dall'algoritmo K-Means classificando a livello minimo tutti i punti critici individuati. L'applicazione dell'algoritmo Isolation Forest funge da ulteriore supervisione rispetto ai risultati emersi dall'applicazione

dell'algoritmo K-Means, generando una stima conservativa rispetto al livello di usura complessivo del veicolo, garantendo una maggiore sicurezza nella classificazione.

Di seguito sono riportati i risultati emersi l'applicazione dell'algoritmo ibrido K-Means (k=5) e Isolation Forest sui dati telemetrici OBD-II del dataset obd2data oggetto di studio. I risultati ottenuti sono stati rappresentati in un diagramma a barre in modo da rendere evidente la distribuzione delle misurazioni rispetto ai 5 livelli di usura definiti su scala 1-5 (1=Critico, 5=Ottimo), scegliendo una colorazione intuitiva dal rosso (alto rischio) al verde (basso rischio). (Figura 13)

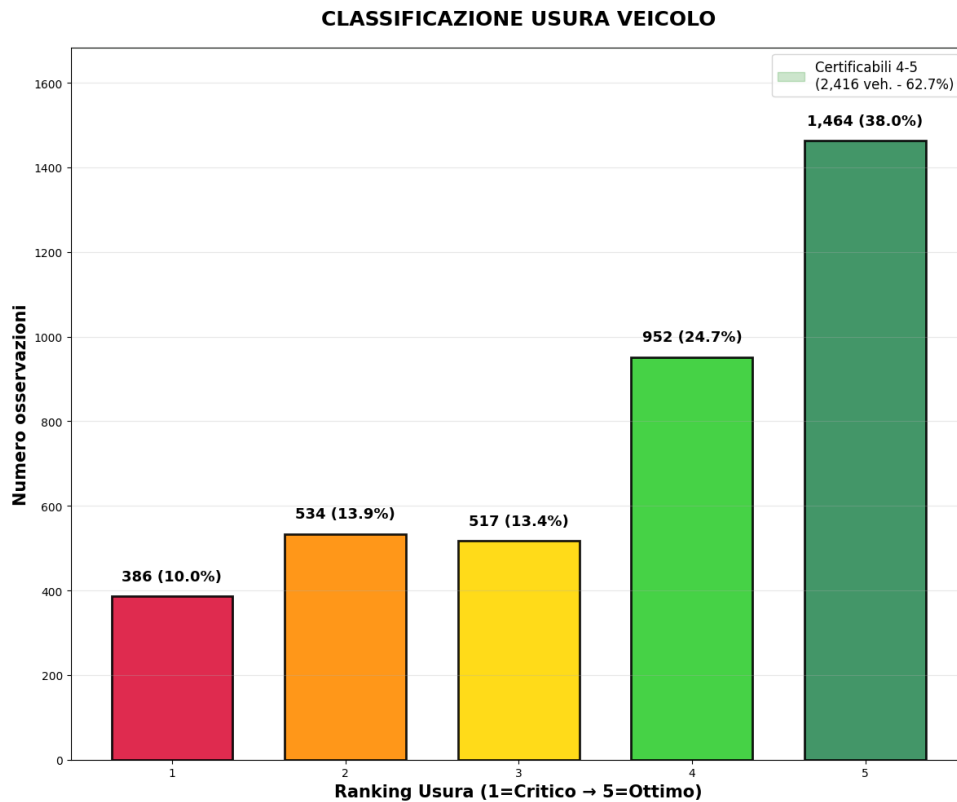


Figura 13: Classificazione usura veicolo sulla base delle misurazioni del dataset analizzato.

Il grafico mostra la distribuzione delle 3853 misurazioni contenute nel dataset obd2data, raccolte durante 10 minuti di guida pendolaristica del veicolo Toyota Etios 2014 (cilindrata 1496 cm³), a seguito dell'applicazione dell'algoritmo ibrido di Machine Learning. I risultati ottenuti dall'analisi mostrano una netta dominanza della barra verde scuro corrispondente al ranking 5, attestando un livello ottimo di usura del veicolo monitorato. In particolare, su 3853 misurazioni totali contenute nel dataset, 1464 sono riconducibili a livelli di stress ottimi, pari al 38%, confermando

un comportamento delle variabili monitorate in linea con una valutazione ottima del livello di usura veicolare.

I ranking 4 e 3, rispettivamente indicati nel grafico con i colori verde chiaro e giallo, registrano rispettivamente 952 misurazioni (24,7%) e 517 misurazioni (13,4%), mentre i livelli 2 e 1 (arancione e rosso) contano 534 (13,9%) e 386 osservazioni (10,0%). Il grafico conferma l'efficacia discriminatoria dell' algoritmo ibrido applicato al dataset, producendo una distribuzione bilanciata e interpretabile che traduce dati telemetrici grezzi in decisioni operative immediate per il mercato dell'usato garantito.

Infatti, dall'analisi delle misurazioni telemetriche riportate nel dataset studiato, emerge che l'area verde relativa ai ranking 4-5 presenta un totale di 2416 misurazioni, rappresentando circa il 63% del totale dei dati campionati, attestando quindi che il veicolo Toyota Etios 2014, può essere rivenduto come usato garantito stimando un livello di usura ottimo e pertanto classificando tale veicolo con ranking 5.

5.5 Analisi dei risultati

Determinare l'accuratezza del metodo proposto per classificare i veicoli rispetto al loro livello di usura motore presenta una sfida intrinseca dovuta al fatto che il dataset utilizzato per lo studio è completamente unsupervised. L'assenza di etichette di riferimento non permette infatti, l'utilizzo di metriche di classificazione tradizionali come l'accuracy o il precision, ma richiede l'utilizzo di metriche di valutazione interna, le quali si basano su un'analisi che permette di determinare quanto i cluster sono ben separati tra loro e compatti al loro interno.

Di seguito si valutano alcune metriche di valutazione interna:

1. Silhouette Coefficient (Coefficiente Silhouette): Questo metodo di analisi, già affrontato nel seguente studio, misura quanto un punto è simile al proprio cluster rispetto agli altri, utilizzando un punteggio che varia da -1 a 1, per cui un valore vicino a 1 indica cluster ben definiti e separati, mentre vicino a -1 indica sovrapposizione.

Partendo dalle metriche di cluster utilizzate, come emerso da analisi precedenti, la classificazione dei dati raccolti nei 5 cluster definiti genera un Silhouette score di 0,235, collocando i risultati dello studio in una fascia "accettabile-buona" per dati telemetrici

reali e perfettamente in linea con i valori riportati in letteratura, per studi condotti su dataset OBD-II analoghi (0.19-0.28) [38][39].

2. Davies-Bouldin Index (Indice Davies-Bouldin): Quest'analisi quantifica la qualità della separazione tra i cluster, calcolando il rapporto tra la dispersione interna al cluster e la distanza tra i cluster, rappresentando una metrica interna complementare al Silhouette score, dove un valore più basso indica una migliore separazione tra i cluster.

L'applicazione di K-Means ($k=5$) produce un indice Davies-Bouldin = 1.669, valore che si colloca nella fascia "adeguata-buona" per clustering su dati telemetrici reali. Questo punteggio indica una separazione cluster sufficientemente distinta nonostante la naturale sovrapposizione tra regimi operativi confinanti (es. Ranking 4 "Buono" vs 3 "Medio"). Il DB=1.669 riflette il compromesso ottimale tra compattezza interna (bassa dispersione entro il ranking) e differenziazione esterna (distanza tra i centri), essenziale per dati multidimensionali eterogenei come i parametri OBD-II. Rispetto ai benchmark settoriali tale risultato si colloca nella fascia "adeguata-buona" (1.5-2.0), confermando la robustezza del metodo applicato, riflettendo un livello di separazione efficace tra i 5 livelli usura [38][39].

3. Inerzia o Somma dei Quadrati Entro-Cluster (WCSS): Questa verifica è utilizzata principalmente nel K-Means, come misura per la compattezza dei cluster, per cui minore è l'inerzia più i punti sono vicini al proprio centroide.

Il valore riscontrato di inerzia intracluster (WCSS) per K-Means $k=5$ risulta pari a 1,787, confermando la scelta metodologica relativa al numero di cluster utilizzati. Tale valore, ottenuto sommando le distanze euclidee quadrate dai centroidi, evidenzia una riduzione del 33% dall'iniziale $k=2$ (WCSS=2,672) con stabilizzazione oltre $k=5$ (-6.6%), riconfermando quanto emerso tramite l'Elbow Method e identificando il "gomito" della curva per $k=5$. Rispetto alla letteratura OBD-II (mediana WCSS=1,856), il risultato posiziona il metodo adottato in un livello alto per qualità partizione ottenuta, garantendo cluster omogenei e stabilmente interpretabili sulla scala usura 1-5 per applicazioni industriali resale [38][39].

4. Calinski-Harabasz Index (Variance Ratio Criterion): Questo indice calcola la qualità della divisione in cluster, massimizzando il rapporto tra varianza tra-cluster e varianza intra-cluster.

L'applicazione di K-Means ($k=5$) produce un Indice Calinski-Harabasz = 522.49, valore che si colloca nella fascia "buona-eccezionale" per dataset multidimensionali reali. Questo punteggio supera, infatti, i benchmark riscontrati in letteratura (mediana CH=456) e conferma eccellente discriminazione naturale tra i 5 ranking di usura veicolare, particolarmente marcata tra estremi (Ranking 5 vs 1). La metrica valida la robustezza decisionale del metodo, garantendo separazione statisticamente solida per certificazioni industriali resale [38][39].

Le metriche di valutazione interna adottate per determinare l'accuratezza dell'analisi condotta, Silhouette Score (0,235), Davies-Bouldin Index (1,669), WCSS (1,787) e Calinski-Harabasz Index (522,49), hanno riscontrato i seguenti valori ottimi, confermando la robustezza del clustering K-Means ($k=5$) su dati OBD-II reali, con risultati classificabili nelle fasce "accettabile-buona" a "buona-eccezionale" rispetto ai benchmark settoriali riscontrati in letteratura [38][39].

L'approccio non supervisionato produce cluster coerenti, separati e compatti, validando la scala 1-5 di usura veicolare adottata per applicazioni operative nel resale usato garantito. La convergenza di tutte e quattro le metriche su $k=5$ dimostra un compromesso ottimale tra qualità partizionale e interpretabilità pratica, rendendo il metodo scientificamente difendibile e immediatamente applicabile in contesti industriali.

CAPITOLO 6

APPLICAZIONI DELLO STUDIO

6.1 Implicazioni sulle decisioni di rivendita

Nel settore della rivendita di auto usate manca un sistema di metriche quantitative oggettive per certificare lo stato di usura dei veicoli, la cui determinazione è attualmente ancorata a paradigmi soggettivi basati su chilometraggio, ispezione visiva e valutazione dello storico delle manutenzioni del veicolo. Le valutazioni fatte per determinare il valore del veicolo che verrà poi reimmesso sul mercato delle auto usate, non prevedono un sistema che determini in maniera oggettiva lo stato di usura dell'auto e il conseguente valore di rivendita del veicolo stesso.

L'applicazione di algoritmi predittivi data-driven, e in particolare l'utilizzo dell'algoritmo ibrido proposto nel seguente studio, consentirebbe di quantificare oggettivamente il livello di usura del veicolo, rappresentando una vera e propria innovazione nel settore della rivendita di auto come usato garantito.

Lo studio condotto propone l'applicazione congiunta dei modelli di Machine Learning K-Means ($k=5$) e Isolation Forest, al fine di analizzare i dati reali raccolti durante l'utilizzo quotidiano dei veicoli per determinare in maniera oggettiva il valore reale dell'auto.

La classificazione generata come output dell'algoritmo ibrido proposto sarà quindi utile in un contesto in cui manca oggettività nell'analisi del livello di usura del veicolo, permettendo una stima reale che mitighi il problema dell'asimmetria informativa in un settore ancora fortemente caratterizzato da incertezza nella valutazione del prodotto e dall'utilizzo di metriche obsolete per la valutazione dello stesso [5].

Questo studio si pone l'obiettivo di risolvere i limiti riscontrati nel mercato dell'usato garantito, fornendo uno strumento solido che consenta di misurare in maniera oggettiva lo stato di usura dell'auto e tradurre questo dato in valore economico del veicolo, calcolato a partire da dati reali raccolti durante l'utilizzo.

6.2 Determinazione del prezzo di rivendita

L'applicazione dell'algoritmo di Machine Learning ibrido, ottenuto dalla combinazione dei modelli predittivi unsupervised K-Means (k=5) e Isolation Forest, genera una classificazione dei veicoli in base al loro livello di usura motore basandosi sulle misurazioni telemetriche raccolte durante le fasi di utilizzo reale della vettura.

Questa classificazione si traduce in decisioni operative utilizzabili per supportare il processo di rivendita dell'usato garantito, risolvendo le asimmetrie informative strutturalmente presenti all'interno del mercato dell'usato. Partendo dalla stima del livello di usura complessivo del veicolo, il risultato ottenuto può essere convertito in valutazioni d'acquisto per i prodotti presenti sul mercato dell'usato, consigliando l'acquisizione di prodotti con misurazioni telemetriche ottimali e suggerendo controlli e interventi di riparazione per quelli che hanno riscontrato livelli di usura critica.

La quantificazione oggettiva del livello di usura complessiva del prodotto, ottenuta tramite l'applicazione di un algoritmo di Machine Learning che sfrutta dati raccolti durante periodi di utilizzo reale del veicolo, garantisce al cliente una maggiore sicurezza sulle prestazioni, rappresentando una maggiore certezza e verifica dello stato di salute e delle condizioni tecniche dei veicoli reimmessi sul mercato dell'usato.

La classificazione delle condizioni di usura delle auto generata dall'applicazione del modello di Machine Learning si traduce in un servizio al cliente che è stato quantificato calcolando il premio aggiuntivo che il cliente è disposto a pagare per ottenere maggiore certezza sulle condizioni complessive del veicolo e per ridurre l'asimmetria informativa tra le parti.

Nella tabella seguente si riassumono le decisioni operative relative a ciascun livello di usura riscontrato attraverso l'applicazione dell'algoritmo, evidenziando i valori del premio aggiuntivo calcolati per ciascun livello di usura veicolare e il conseguente prezzo di rivendita proposto, i quali risultati sono discussi di seguito (Tabella 12) [41][42][43][44].

Tabella 12: Classificazione dei livelli di usura veicolare e stima del prezzo di rivendita.

Ranking	Valore usura	Significato	Premio	Prezzo
5	Ottimo	Acquisto consigliato	+ 24%	€18.600
4	Buono	Veicolo affidabile	+ 12%	€16.800
3	Normale	Funzionante	+ 1 %	€15.150
2	Elevato	Riparazioni necessarie		
1	Critico	Acquisto non consigliato		

Per determinare il prezzo complessivo al quale rivendere l'auto sul mercato dell'usato garantito, tenendo conto della maggiore garanzia che viene offerta al cliente, si assume un valore base di mercato pari a €15.000, che rappresenta la media nazionale 2025 per veicoli del segmento C con età compresa tra 3-5 anni e chilometraggio medio di circa 80.000km [41]. A tale valore si dovrà sommare il premio economico associato alla riduzione di asimmetria informativa dovuta all'applicazione di un metodo oggettivo di stima del livello di usura veicolare. Nel dettaglio, il premio aggiuntivo calcolato per ciascun livello di usura è stato determinato secondo un approccio conservativo basato su benchmark settoriali dei programmi di rivendita dell'usato garantito delle principali case automobilistiche.

Innanzitutto, si è determinato il premio per il ranking 5 (ottimo) che identifica il maggior livello di qualità del veicolo che si può riscontrare dall'applicazione dell'algoritmo. Tale premio economico è stato determinato in base alla media dei programmi di rivendita di veicoli certificati delle principali case automobilistiche premium, tra cui BMW Premium Selection, che prevede un incremento del 26%, Mercedes-Benz Certified del 24% e Audi Prima Scelta del 23%, per una media settoriale del 24,3% e arrotondata conservativamente a 24% [42][43][44].

Successivamente si è determinato il premio da associare al ranking 3 (normale), identificando una condizione di usura media coerente con un veicolo con prestazioni standard nel mercato dell'usato. Per tale ragione il premio previsto per questo livello di qualità del veicolo è minimo dell'1%.

Sulla base dei premi identificati per il livello 5 (ottimo) e per il livello 3 (normale), si è potuto determinare il premio relativo al livello di usura 4 (buono), calcolato come classe intermedia tra le due condizioni citate, ottenendo un premio proporzionale da attribuire a tale classe, pari al 12%, ottenuto come valore conservativo medio tra il 24% del ranking 5 e l'1% del ranking 3.

L'applicazione dei premi determinati al valore base di mercato considerato per un veicolo di segmento C definito precedentemente, pari a €15.000 produce per il ranking 5 un prezzo di rivendita di €18.600 a fronte di un premio del 24%, per il ranking 4 si prevede un prezzo di rivendita di €16.800 per un premio del 12% e infine un prezzo di €15.150 per veicoli riconducibili al ranking 3, per i quali si considera un premio del 1%.

I ranking 1 e 2 restano esclusi da questa analisi, in quanto non prevedono premi economici poiché per livelli di usura così severi si sconsiglia l'acquisto, in particolare per il ranking 1 (critico)

dove il veicolo non è idoneo alla rivendita, mentre per il ranking 2 (elevato) si raccomanda una revisione completa prima di qualsiasi decisione commerciale.

6.3 Garanzia dell'usato

Analogamente all'analisi condotta per determinare i prezzi dell'usato garantito, sulla base dei risultati emersi dalla valutazione dell'usura veicolare condotta attraverso l'applicazione dell'algoritmo di Machine Learning proposto, si possono definire diversi livelli di garanzia associati a ciascun ranking di usura del veicolo. Nella tabella seguente si riassumono i livelli di garanzia relativi ai ranking di usura veicolare determinati dall'algoritmo (Tabella 13) [43][44][45][46][47].

Tabella 13: Classificazione dei livelli di usura veicolare e stima del livello di garanzia.

Ranking	Valore usura	Significato	Garanzia usato
5	Ottimo	Acquisto consigliato	36 mesi
4	Buono	Veicolo affidabile	24 mesi
3	Normale	Funzionante	12 mesi
2	Elevato	Riparazioni necessarie	
1	Critico	Acquisto non consigliato	

La durata della garanzia per ciascun livello di usura veicolare è stata determinata a partire da un'analisi di benchmark condotta sui principali programmi certificati di rivendita dell'usato garantito delle maggiori case automobilistiche.

In particolare, per il ranking 5 (ottimo) si adotta una garanzia di 36 mesi, valore conservativo derivato dalla media delle coperture offerte da BMW Premium Selection (48 mesi), Mercedes-Benz Certified (12-48 mesi), Porsche Approved (36 mesi) e Volvo Select (36 mesi) [43][44][45][46].

Per il ranking 3 (normale), corrispondente a una condizione di usura media standard di mercato, si prevede il minimo legale di 12 mesi stabilito dal Codice del Consumo (D.Lgs. 206/2005) per le vendite di beni di seconda mano tra professionisti e consumatori [47].

Infine, la garanzia relativa al ranking 4 (buono), è stato calcolato a partire dalle durate delle garanzie che si sono determinate per il livello 5 (ottimo) e livello 3 (normale), in quanto classe intermedia tra i ranking due ranking, determinando una garanzia di 24 mesi.

6.4 Impatto economico sul mercato dell'usato garantito

L'introduzione del metodo predittivo *unsupervised* per la classificazione usura veicolare mediante ranking 1-5 genera un impatto significativo sul mercato dell'usato garantito, introducendo maggiore trasparenza e risolvendo il problema strutturale dell'asimmetria informativa tra rivenditori e acquirenti che ha sempre penalizzato il settore della rivendita delle auto [5]. L'algoritmo ibrido applicato ai dati di utilizzo reali trasferisce questa informazione nascosta ai consumatori attraverso una classificazione trasparente e verificabile, trasformando l'incertezza percepita in fiducia misurabile e accrescendo la loro disponibilità a pagare prezzi più alti per i veicoli con ranking 4-5, mentre i ranking inferiori vengono naturalmente scremati dal circuito commerciale. Questo riequilibrio informativo attiva un circolo virtuoso che si traduce in un aumento del market share, in quanto l'applicazione dell'algoritmo di Machine Learning aiuterebbe ad attrarre maggiore domanda di mercato grazie alla credibilità oggettiva del servizio, amplificano i volumi di vendita del segmento dell'usato garantito [48]. Di seguito si analizzano nel dettaglio le due leve di crescita economica del settore che sono state citate.

6.4.1 Aumento del margine dei rivenditori

Il mercato secondario dei beni durevoli è caratterizzato dalla presenza di asimmetria informativa, ossia una situazione in cui una delle parti possiede più informazioni rispetto all'altra e il quale effetto è stato esaminato dall'economista americano George Arthur Akerlof nel suo articolo pubblicato sul *Quarterly Journal of Economics* dal titolo "The Market for Lemons: Quality Uncertainty and the Market Mechanism". Akerlof spiega come la presenza di asimmetria informativa nel mercato delle auto usate porti il consumatore, non disponendo di sufficienti informazioni sulle condizioni di usura delle auto, a valutare i prodotti come se fossero di media qualità [5]. Lo scenario descritto tende a sfavorire le auto di alta qualità presenti sul mercato, per le quali i consumatori non saranno disposti a pagare un prezzo che rispecchi appieno il livello del prodotto, in quanto non hanno abbastanza informazioni per poter valutare e riconoscerne il valore di quell'auto rispetto alle altre.

In questo senso l'applicazione dell'algoritmo proposto in questo studio risolve il problema dell'asimmetria informativa, fornendo un metodo oggettivo che consente di verificare la qualità dell'auto rivenduta come usato certificato. La classificazione del veicolo rispetto al suo livello di usura permette ai clienti di verificare la qualità dell'auto, cambiano le dinamiche del settore. Il

cliente, infatti, avendo prova del valore reale dell'auto sarà disposto a pagare un prezzo maggiore per un'auto in ottimo stato, in quanto può contare su un metodo che si basa su dati reali raccolti durante l'utilizzo della vettura.

Sulla base delle considerazioni fatte nei in precedenza, il rivenditore può aumentare il prezzo di rivendita del veicolo usato giustificando tale aumento con la riduzione di asimmetria informativa e la garanzia di una valutazione oggettiva dello stato di usura dell'auto. Questo permette al rivenditore di guadagnare un premio sul prezzo base del 24% per auto in ottimo stato (ranking 5), del 12% per vetture che presentano un livello di usura buono (Ranking 4) e soltanto dell'1% per auto classificate con livello di usura normale (Ranking 3), rimando quindi in media un aumento del margine lordo per il rivenditore circa del 15,6% sul prezzo base.

6.4.2 Aumento del market share

L'introduzione di un algoritmo di Machine Learning per la valutazione predittiva dell'usura veicolare basato sull'elaborazione di dati telemetrici OBD-II, può ridurre significativamente l'asimmetria informativa nel mercato dell'usato garantito, con impatti quantificabili sui volumi di vendita e sul valore di mercato dei veicoli certificati.

Courty e Pagliano (2018), analizzando come la trasparenza informativa, derivante dalla possibilità di avere un feedback informativo sulla qualità dei veicoli usati attraverso dati diagnostici, incrementi i volumi di vendita delle vetture certificate del 52% entro 18 mesi dall'implementazione della soluzione di telemetria, trasformando il mercato della rivendita delle auto usate da mercato "avvelenato dai limoni" verso una condizione di maggior equilibrio competitivo e maggiore offerta di veicoli premium [5][48].

Applicando tali evidenze al contesto italiano, dove il mercato dell'usato automobilistico genera circa 27 miliardi di euro annui e il segmento garantito rappresenta il 15-20% (4,05-5,4 miliardi €), la soluzione di Machine Learning proposta produrrebbe un aumento pari a +2,1-2,8 miliardi di euro, considerando un incremento dei volumi del 52% come previsto dagli studi presenti in letteratura [48][49].

In conclusione, l'applicazione sistematica del modello predittivo data-driven sull'intero mercato italiano di auto usate produrrebbe un impatto economico quantificabile in circa €2,5 miliardi di valore economico annuo per i rivenditori [49].

6.5 Sostenibilità

In ultima analisi, l'applicazione pratica del modello di classificazione unsupervised basato su K-Means e Isolation Forest, genera implicazioni significative per la sostenibilità, non solo dal punto di vista sociale (trasparenza mercato usato), ma anche sul piano ambientale.

La certificazione oggettiva dell'usura di un veicolo usato (ranking 1-5), consente di estendere significativamente la vita media utile del parco circolante, posticipando demolizioni premature e produzione nuovi veicoli. Questo aspetto, dal punto di vista ambientale, consente di evitare emissioni derivanti dalla produzione di un nuovo veicolo sostitutivo, evitando l'emissione di circa 5,6-7,5 tonnellate di CO₂ per la realizzazione di un'auto benzina o diesel, oltre al risparmio di materie prime critiche come acciaio, alluminio e plastiche [51].

Inoltre, assumendo che l'introduzione di questo modello come strumento di valutazione della qualità dell'auto usata generi un aumento dei volumi di vendita del 52%, come emerso da studi condotti in letteratura, l'applicazione dell'algoritmo proposto porterebbe ad una riduzione di emissioni di CO₂ in Italia di circa 1,75-2,34 milioni di tonnellate annue [48][49].

L'impatto ambientale dell'algoritmo si quantifica considerando il mercato certificato italiano conta attualmente 500.000-650.000 veicoli annui (10-15% dei 5,5 milioni passaggi proprietà totali), per un valore di 4-5,4 miliardi € su totale 27 miliardi € del second-hand automotive [41]. L'espansione del 52% genera 260.000-338.000 transazioni aggiuntive di veicoli usati certificati, evitando altrettante produzioni ex-novo con emissione media di 5,6-7,5 tonnellate di CO₂ per unità. Il risultato è una riduzione di 1,46-2,54 milioni di tonnellate CO₂ annue, equivalente all'1,5-2,5% delle emissioni totali del settore trasporti stradali italiani (100 milioni tonnellate CO₂/anno) [52].

L'innovazione introdotta dall'applicazione del modello di Machine Learning per la stima del livello di usura dei veicoli non rappresenterebbe quindi soltanto una soluzione per stimare in modo oggettivo la qualità dell'auto in un contesto fortemente interessato dalla presenza di informazione asimmetrica tra la parti, ma contribuirebbe anche ad allineare l'operato nazionale rispetto agli European Green Deal, contribuendo alla riduzione delle emissioni e alla transizione digitale [50].

CONCLUSIONI

La presente tesi ha sviluppato e validato un modello predittivo ibrido data-driven basato su K-Means clustering ($k=5$) e Isolation Forest applicato a 3853 misurazioni telemetriche OBD-II raccolti su un veicolo Toyota Etios 2014 (cilindrata 1496 cm³) in condizioni di guida reale per una durata di 10 minuti ciascuna [29]. L'applicazione dell'algoritmo ibrido di Machine Learning produce una classificazione oggettiva dello stato di usura del veicolo analizzato attraverso cinque ranking (1-5), sostituendo definitivamente i paradigmi soggettivi tradizionali basati su chilometraggio, ispezione visiva e storico manutenzioni con un metodo basato su algoritmi di Machine Learning che sfrutta dati reali, raccolti dal funzionamento operativo della vettura, eliminando l'asimmetria informativa strutturale del mercato delle auto usate [5].

L'applicazione della soluzione proposta genera un significativo impatto sul mercato della rivendita di auto usate, stimato in un aumento del valore del mercato di €2,5 miliardi, a fronte di un aumento di market share del 52%, giustificato dall'utilizzo di una metodologia che assicura la qualità dell'acquisto al cliente [48].

L'impatto di questo studio non si limita soltanto ad una crescita economica del settore dell'usato garantito, ma ha anche un'importante valenza in ambito ambientale. La stima del livello di usura dell'auto contribuisce ad attestarne in maniera oggettiva il livello di qualità, fornendo maggiore sicurezza al cliente relativamente all'acquisto di un'auto usata. La certezza offerta al cliente va ad incidere sulle sue decisioni d'acquisto, riflettendosi dal punto di vista ambientale in una scelta sostenibile del consumatore che in questo modo contribuisce ad evitare la produzione di emissioni derivanti dalla creazione di nuove vetture, stimando una riduzione di emissioni di CO₂ in Italia di circa 1,75-2,34 milioni di tonnellate annue.

Il modello proposto non rappresenta un'ottimizzazione incrementale ma un'innovazione disruptive che ridefinisce le dinamiche interne del settore della mobilità usata, dimostrando che l'intelligenza artificiale applicata alla telemetria veicolare può essere assumere una posizione determinante in tematiche centrali nel XXI secolo come la sostenibilità ambientale attraverso una valutazione oggettiva del prodotto reimmesso sul mercato.

La telemetrica OBD-II, precedentemente sottoutilizzata, diventa leva strategica per la sostenibilità della mobilità, conferendo validazione e credibilità in un settore fino ad ora fortemente caratterizzato da incertezza, supportando i consumatori nel fare scelte di acquisto

sostenibili, grazie alla certificazione dell'usato sicuro supportato da valutazioni realistiche e oggettive.

BIBLIOGRAFIA

- [1] Wadhvani P. Jaiswal S. (2025). Market Size by Vehicle. Report ID: GMI5019. Consulted: Nov. 7, 2025.
- [2] Cinieri S. (2019). Predictive maintenance of industrial vehicles based on supervised machine learning techniques. [Tesi di Laurea Magistrale, Ingegneria Informatica]. Politecnico di Torino. Consulted: Nov. 9, 2025. Available: <https://webthesis.biblio.polito.it/id/eprint/13148>
- [3] Rahim, M. A., Rahman, M. A., Rahman, M. M., Asyhari, A. T., Bhuiyan, M. Z. A., & Ramasamy, D. (2021). Evolution of IoT-enabled connectivity and applications in automotive industry: A review. *Vehicular communications*, 27, 100285. Consulted: Nov. 10, 2025. Available: <https://doi.org/10.1016/j.vehcom.2020.100285>
- [4] Michailidis, E. T., Panagiotopoulou, A., & Papadakis, A. (2025). A review of OBD-II-based machine learning applications for sustainable, efficient, secure, and safe vehicle driving. *Sensors*, 25(13), 4057. Consulted: Nov. 26, 2025. Available: <https://doi.org/10.3390/s25134057>
- [5] Akerlof, G. A. (1978). The market for “lemons”: Quality uncertainty and the market mechanism. In *Uncertainty in economics* (pp. 235-251). Academic Press. Consulted: Nov. 9, 2025. Available: <https://doi.org/10.1016/B978-0-12-214850-7.50022-X>Get rights and content
- [6] Orsato, R. J., & Wells, P. (2007). U-turn: the rise and demise of the automobile industry. *Journal of Cleaner Production*, 15(11-12), 994-1006. Consulted: Feb. 11, 2025. Available: <https://doi.org/10.1016/j.jclepro.2006.05.019>
- [7] Nanayakkara, S., & Meegama, R. G. N. (2024). A deep learning approach for scoring quality in vehicle grading. *Engineering Applications of Artificial Intelligence*, 131, 107812. Consulted: Feb. 21, 2025. Available: <https://doi.org/10.1016/j.engappai.2023.107812>
- [8] Nissan USA. Certified pre-owned inspection checklist. Consulted Feb. 21, 2025. Available: <https://www.nissanusa.com/content/dam/Nissan/us/shopping-tools/certified-pre-owned/167%20CPO%20Checklist.pdf>
- [9] Treccani. Telemetria. Vocabolario Treccani. Istituto della Enciclopedia Italiana. Consulted: Gen. 15, 2025. Available: <https://www.treccani.it/vocabolario/telemetria/>

- [10] Fundamental Business Insights (2025). Automotive Embedded Telematics Market Size, Share & Industry Analysis. Report ID: FBI4440 Consulted: Nov. 19, 2025. Available: <https://www.fundamentalbusinessinsights.com/industry-report/automotive-embedded-telematics-market-4440>
- [11] Benevento L. (2023). Trasformazione digitale dell'auto e uno studio preliminare sui need legati al veicolo autonomo [Tesi di Laurea Magistrale, Ingegneria Gestionale]. Politecnico di Torino. Consulted: Ott. 30, 2025. Available: <https://webthesis.biblio.polito.it/29706/1/tesi.pdf>
- [12] De Maria G. (2022). Panoramica dei principali test nell'ambito automotive [Tesi di Laurea Triennale, Ingegneria Meccatronica]. Università degli studi di Padova. Consulted: Ott. 25, 2025. Available: https://thesis.unipd.it/retrieve/27952e57-a053-4033-86d7-70fb74a6fc0a/DeMaria_Giuseppe.pdf
- [13] Morari M. (2022). I sensori nelle automobili: tecnologia e applicazioni [Tesi di Laurea Triennale, Ingegneria Biomedica]. Università degli Studi di Padova. Consulted: Dec. 7, 2025. Available: https://thesis.unipd.it/retrieve/0a0652b9-f6f1-4536-9610-38232cb30266/Morari_Michele.pdf
- [14] Hengrui L. (2022). Battery state of health and state of charge estimation with related application [Tesi di Laurea Magistrale, Automotive Engineering]. Politecnico di Torino. Consulted: Gen. 9, 2026. Available: <https://webthesis.biblio.polito.it/24360/1/tesi.pdf>
- [15] Falletto P. (2025). Studio sperimentale di attrito, usura e comportamento termico mediante banco prova twin-disc in ambito ferroviario [Tesi di Laurea Magistrale, Ingegneria Meccanica]. Politecnico di Torino. Consulted: Gen. 14, 2026. Available: <https://webthesis.biblio.polito.it/37544/1/tesi.pdf>
- [16] Marotta F. (2010). Analisi sperimentale per la valutazione dell'intensità di detonazione al variare del numero di ottano del carburante [Tesi di Laurea Magistrale, Ingegneria Industriale]. Politecnico di Milano. Consulted: Gen. 15, 2026. Available: <https://www.politesi.polimi.it/retrieve/a81cb059-92a3-616b-e053-1605fe0a889a/Tesi%20di%20Laurea%20di%20Fabio%20Marotta.pdf>
- [17] Parlamento europeo e Consiglio dell'Unione europea (2015). Regolamento (UE) 2015/758 del 29 aprile 2015 relativo ai requisiti di omologazione per l'introduzione del sistema eCall basato sul numero di emergenza 112 e che modifica la direttiva 2007/46/CE. Gazzetta

ufficiale dell'Unione europea. Consulted: Nov. 21, 2025. Available: <https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:32015R0758&from=ET>

[18] Parlamento europeo e Consiglio dell'Unione europea (2019). Regolamento (UE) 2019/2144 del 27 novembre 2019 relativo ai requisiti di omologazione dei veicoli a motore e dei loro rimorchi, nonché di sistemi, componenti ed entità tecniche destinati a tali veicoli per quanto riguarda la sicurezza generale e la protezione degli occupanti dei veicoli e degli altri utenti vulnerabili della strada. Gazzetta ufficiale dell'Unione europea. Consulted: Nov. 21, 2025. Available: <https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:32019R2144>

[19] Parlamento europeo e Consiglio dell'Unione europea (2025). Regolamento (UE) 2025/1329 dell'11 luglio 2025. Gazzetta ufficiale dell'Unione europea. Consulted: Nov. 21, 2025. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202501329

[20] Parlamento europeo e Consiglio dell'Unione europea. (2018). Regolamento (UE) 2018/1947. Gazzetta ufficiale dell'Unione europea. Consulted: Nov. 21, 2025. Available: <https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:42018X1947&from=EN>

[21] European Commission. (2016). *Commission Regulation (EU) 2016/646 of 20 April 2016 amending Regulation (EC) No 692/2008 as regards emissions from light passenger and commercial vehicles (Euro 6)* (Text with EEA relevance). *Official Journal of the European Union*, L 103, pp. 1-18. Consulted: Nov. 23, 2025. Available: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0646>

[22] United Nations Economic Commission for Europe. (2020). *Regulation No. 83: Uniform provisions concerning the approval of passenger cars and their components with regard to the emission of gaseous pollutants (latest amendment incorporating OBD requirements)*. Consulted: Nov. 23, 2025. Available: https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=OJ:L_202401312

[23] European Parliament & Council of the European Union. (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing*

[24] United Nations Economic Commission for Europe. (2007). *Regolamento n. 83 della Commissione economica per l'Europa delle Nazioni Unite (UN/ECE) — Disposizioni uniformi relative*

all'omologazione dei veicoli a motore per quanto riguarda le emissioni di inquinanti gassosi. Gazzetta Ufficiale dell'Unione Europea, L 172 (3 luglio 2015). Consulted: Feb. 23, 2026. Available: <https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=OJ:L:2015:172:FULL>

[25] Denton, T. (2020). *Advanced automotive fault diagnosis: automotive technology: vehicle maintenance and repair*. Routledge. Consulted: Dec. 02, 2025. Available: https://e-tk.lntu.edu.ua/pluginfile.php/17464/mod_resource/content/1/denton_t_advanced_automotive_fault_diagnosis-1-31.pdf

[26] Fleming, W. J. (2002). Overview of automotive sensors. *IEEE sensors journal*, 1(4), 296-308. Consulted: Dec: 02, 2025. Available: https://ieeexplore.ieee.org/iel5/7361/21198/00983469.pdf?casa_token=44Ks7OewqPAAA:AAA:EXDvE0Xy2O29Vq_CN_K_w1VfmDEMQ4wEn5JqlENHhBoTkY53vTV1SeDcaIfUP43D_oiNik

[27] ISO, *ISO 15765-2*. Consulted: Dec. 09, 2025. Available: <https://www.iso.org/standard/66574.html>

[28] SAE International, *J1979-DA, Digital Annex of E/E Diagnostic Test Modes*. Consulted: Dec. 9, 2025. Available: https://www.sae.org/standards/j1979da_201905-j1979-da-digital-annex-e-e-diagnostic-test-modes

[29] Eron93br (2023). *OBD2Data: Toyota Etios 2014 OBD-II telemetry dataset*. Kaggle. Consulted: Dec. 29, 2025. Available: <https://www.kaggle.com/datasets/eron93br/obd2data>

[30] Istat (2025), *Spostamenti quotidiani e nuove forme di mobilità*. Consulted: Gen. 30, 2026. Available: <https://www.istat.it/wp-content/uploads/2025/12/PSCL-2025-Istat-sedi-Roma.pdf>

[31] Kumar, A., et al. (2025). Predictive Maintenance of Automotive Engines Using Machine Learning and Sensor Data Fusion. *International Journal of Innovative Research in Technology (IJIRT)*, 179974. Consulted: Mar. 02, 2026. Available: https://ijirt.org/publishedpaper/IJIRT179974_PAPER.pdf

[32] Xu, W., Le Roux de Bretagne, O., Grasso, M., Harrison, M. F., Carless, O., & StLeger-Harris, C. (2024). On a numerical methodology to assess the fatigue life of connecting rods. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile*. Consulted: Feb. 02, 2025. Available: <https://journals.sagepub.com/doi/10.1177/09544070231180996>

- [33] Hofmockel, J., & Sax, E. (2018). Isolation Forest for Anomaly Detection in Raw Vehicle Sensor Data. In *VEHITS* (pp. 411-416). Consulted: Gen. 18, 2025. Available: <https://pdfs.semanticscholar.org/251e/87bcd1c12ce4e54d3facdd8c34fe98ac1ae8.pdf>
- [34] Menard, S. (2001). *Applied logistic regression analysis*. SAGE publications. Consulted: Mar. 07, 2026. Available: [https://books.google.com/books?hl=it&lr=&id=JbVIDwAAQBAJ&oi=fnd&pg=PP7&dq=Menard,+S.+\(1995\).+Applied+Logistic+Regression+Analysis.+Sage+Publications.&ots=4Qa127A0t_&sig=Z-NRZspYaVCzXk7C_zwFJJ0z54g](https://books.google.com/books?hl=it&lr=&id=JbVIDwAAQBAJ&oi=fnd&pg=PP7&dq=Menard,+S.+(1995).+Applied+Logistic+Regression+Analysis.+Sage+Publications.&ots=4Qa127A0t_&sig=Z-NRZspYaVCzXk7C_zwFJJ0z54g)
- [35] Nickerson, A., Japkowicz, N., & Milos, E. E. (2001, January). Using unsupervised learning to guide resampling in imbalanced data sets. In *International workshop on artificial intelligence and statistics* (pp. 224-228). PMLR. Consulted: Mar. 09, 2026. Available: <http://proceedings.mlr.press/r3/nickerson01a/nickerson01a.pdf>
- [36] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3), 1-58. Consulted: Mar. 09, 2026. Available: <https://dl.acm.org/doi/pdf/10.1145/1541880.1541882>
- [37] He, T., Zhang, L., Kong, F., & Salekin, A. (2020, July). Exploring inherent sensor redundancy for automotive anomaly detection. In *2020 57th acm/ieee design automation conference (dac)* (pp. 1-6). IEEE. Consulted: Gen. 22, 2026. Available: https://asalekin.github.io/doc/DAC2020_anomaly_detection.pdf
- [38] Shi, X., Wong, Y. D., Chai, C., Li, M. Z. F., Chen, T., & Zeng, Z. (2022). Automatic clustering for unsupervised risk diagnosis of vehicle driving for smart road. *IEEE Transactions on Intelligent Transportation Systems*, 23(10), 17451-17465. Consulted: Gen. 22, 2026. Available: <https://arxiv.org/pdf/2011.11933>
- [39] Hofmockel, J., & Sax, E. (2018). Isolation Forest for Anomaly Detection in Raw Vehicle Sensor Data. In *VEHITS* (pp. 411-416). Consulted: Gen. 22, 2026, Available: <https://pdfs.semanticscholar.org/251e/87bcd1c12ce4e54d3facdd8c34fe98ac1ae8.pdf>
- [40] Elmahalwy, A. M., Mousa, H. M., & Amin, K. M. (2023). New hybrid ensemble method for anomaly detection in data science. *International Journal of Electrical and Computer Engineering*, 13(3), 3498-3508. Consulted: Gen. 22, 2026. Available: https://www.academia.edu/download/99304136/108_29157_EMr_N.pdf

- [41] ACI. (2025). Annuario statistico 2025: Tutti i numeri delle auto in Italia. Consulted. Mar. 02, 2026. Available: <https://aci.gov.it/comunicati-stampa/annuario-statistico-2025-tutti-i-numeri-delle-auto-in-italia/>
- [42] Audi Italia. (2025). Audi Prima Scelta. Consulted: Mar. 02, 2026. Available: <https://www.audi.it/it/usato/>
- [43] BMW Italia. (2025). BMW Premium Selection. Consulted: Mar. 02, 2026. Available: <https://www.bmw.it/it/topics/offerte-e-servizi/bmw-premium-selection.html>
- [44] Mercedes-Benz Italia. (2025). Mercedes-Benz Certified. Consulted: Mar. 02, 2026. Available: <https://www.mercedes-benz.it/passengercars/services/used.html>
- [45] Porsche Italia. (2025). Porsche Approved: Garanzia pre-owned. Consulted. Mar. 02, 2026. Available: <https://www.porsche.com/italy/it/approved/>
- [46] Volvo Car Italia. (2025). Volvo Selekt: Usato certificato Volvo. Consulted: Mar. 02, 2026. Available: <https://www.volvocars.com/it/cars/used-cars/volvo-selekt/>
- [47] Italia. (2005). Decreto Legislativo 6 settembre 2005, n. 206. Codice del Consumo. Gazzetta Ufficiale n. 261 del 8 novembre 2005. Consulted: Mar. 02, 2026. Available: <http://www.gazzettaufficiale.it/eli/id/2005/10/18/05G0233/sg>
- [48] Courty, P., & Pagliano, S. (2018). Does information feedback affect quality in the used car market? (NBER Working Paper No. 24889). National Bureau of Economic Research. Consulted: Mar. 04, 2026. Available: https://www.nber.org/system/files/working_papers/w24889/w24889.pdf
- [49] Osservatorio Second Hand Economy. (2025, 15 maggio). Auto usata, mercato in forte crescita: raggiunti 27 miliardi di euro di valore generato. La Repubblica. Consulted: Mar. 04, 2026. Available: https://finanza.repubblica.it/News/2025/05/15/auto_usata_mercato_in_forte_crescita
- [50] Commissione Europea. (2019). The European Green Deal (COM(2019) 640 final). Consulted: Mar. 04, 2026. Available: <https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:52019DC0640>

[51] Transport & Environment. (2021). How clean are electric cars? T&E Briefing. Consulted: Mar. 04, 2026. Available:

https://www.transportenvironment.org/sites/te/files/publications/2021_12_How_clean_are_electric_cars_211201.pdf

[52] ISPRA, "Emissioni di gas serra dai trasporti," Indicatori Ambientali, Roma, Italia, 2025. Consulted. Mar. 02, 2026. Available:

<https://indicatoriambientali.isprambiente.it/it/trasporti/emissioni-di-gas-serra-dai-trasporti>.

