**Car Recommendation System**

Hamza Jabari and Elham Derbas

1 Al Hussein Technical University, Amman 11831, Jordan

21110127@htu.edu.jo, elham.derbas@htu.edu.jo

**Abstract**

**Keywords:** Artificial Intelligence, Recommendation system, Fuzzy Logic, automative AI, Decision support systems, vehicle selection, AI in transportation, User-centric recommendation, Data-Driven decision making, K-means, web-based recommendations, HTML.

This paper presents the design, development, and evaluation of an intelligent car recommendation system that integrates artificial intelligence techniques with fuzzy logic to manage uncertainty in user preferences. By leveraging a mixed-methods approach, the system combines primary data collected from user surveys with secondary market data sourced from a comprehensive Kaggle automotive dataset. The recommendation engine utilizes fuzzy logic to interpret and process vague user inputs (such as “affordable” or “good condition”) and applies K-means clustering to segment vehicles based on key market attributes including price, age, and condition. The hybrid approach not only improves the personalization of recommendations but also enhances the system’s scalability and responsiveness. A Flask-based web interface is implemented to deliver real-time recommendations, and performance is evaluated through both quantitative metrics and user feedback. The findings demonstrate that the integration of fuzzy logic with clustering algorithms can significantly bridge the gap between subjective user expectations and objective market trends, paving the way for more accurate and user-centric automotive decision support systems. Future work will explore the incorporation of dynamic, real-time data streams and further refinement of the fuzzy rule base to better capture evolving consumer behavior.

Table of Contents

[Introduction 3](#_Toc189592648)

[The research questions driving this work are: 3](#_Toc189592649)

[Research Papers Titles: 4](#_Toc189592650)

[Ibrahim et al. (2023): Hybrid RBM and WaterWheel Plant Algorithm for Electric Vehicle Recommendation Systems 5](#_Toc189592651)

[Nawara & Kashef (2021): IoT-Based Context-Aware Recommendations (IoT-CARS) 6](#_Toc189592652)

[Zheng et al. (2016): M-Distance for Collaborative Filtering 7](#_Toc189592653)

[Bodduluri et al. (2024): Hybrid Recommendation Systems in E-Commerce – A Systematic Review 9](#_Toc189592654)

[Maruthavani & Shantharajah (2024): Distributed Healthcare Recommendations Using Apache Spark and Fuzzy Neural Networks 10](#_Toc189592655)

[Wang et al. (2017): Leveraging Natural Language Processing for Hybrid Car Recommendations 12](#_Toc189592656)

[Synthesis and Research Gaps 14](#_Toc189592657)

[Data collection and Description 14](#_Toc189592658)

[Primary Data: Survey Results and Analysis 14](#_Toc189592659)

[Secondary Data: Kaggle Automotive Dataset 18](#_Toc189592660)

[Research Approach and Methodologies 20](#_Toc189592661)

[Onion Model 20](#_Toc189592662)

[Research Methodology 22](#_Toc189592663)

[Results and Discussion 25](#_Toc189592664)

[Conclusion and Recommendations 28](#_Toc189592665)

[References 31](#_Toc189592666)

# Introduction

**Artificial Intelligence (AI)** transformed multiple industries by allowing systems to simulate human intelligence, make informed decisions, and adapt to dynamic environments. In terms of recommendation systems, AI has proven to be a game-changer through providing personalized and efficient solutions adapted to individual needs. Whilst the complexity of user preferences increases, usual methods struggle to handle uncertainty and imprecision, this is where advanced techniques like fuzzy logic are discussed.

This paper focuses on creating an intelligent car recommendation system that influences AI and fuzzy logic to handle the uncertainties in user preferences. Disparate the conventional recommendation systems, that rely on fixed algorithms, the proposed system implements fuzzy logic to accommodate nuanced user-defined specifications, for instance vehicle condition, price, and personal priorities. By addressing these complexities, the system aims to ensure accurate, scalable, and user-centric recommendations while adapting to diverse and dynamic inputs.

**Objective:**

The objective of this research is to develop a car recommendation system that can suggest cars to potential buyers based on their preferences and past purchase behavior. The system will leverage the historical sales data to identify patterns and similarities between different cars and user preferences.

## The research questions driving this work are:

1. **Integration of Fuzzy Logic:**  
   How can fuzzy logic be effectively integrated with machine learning techniques (such as K-means clustering) to manage the inherent uncertainty in user preferences within car recommendation systems?
2. **Impact of Data Fusion:**  
   To what extent does combining primary survey data with secondary market data improve the accuracy and personalization of car recommendations compared to using a single data source?
3. **Evaluation of Clustering Techniques:**  
   What is the effect of K-means clustering on segmenting automotive market data, and how does this segmentation influence the quality of personalized recommendations?
4. **Optimization of Fuzzy Rules:**  
   How do different configurations and complexities of fuzzy logic rule sets impact the system’s performance in terms of computational efficiency and recommendation accuracy?
5. **Scalability and Real-Time Adaptability:**  
   What challenges are associated with scaling the proposed recommendation system for real-time applications, and how can a hybrid fuzzy logic approach mitigate these challenges while maintaining system responsiveness and accuracy?

# Research Papers Titles:

1. [Unscrambling Customer Recommendations: A Novel LSTM Ensemble Approach in Airline Recommendation Prediction Using Online Reviews | IEEE Journals & Magazine | IEEE Xplore](https://ieeexplore.ieee.org/document/9875228)
2. [Personalized Electricity Tariff Recommendation Method for Residential Customers Lacking Historical Metering Data Incorporating Customer Profiles and Behavioral Changes | IEEE Journals & Magazine | IEEE Xplore](https://ieeexplore.ieee.org/document/10520307)
3. [OntoCommerce: Incorporating Ontology and Sequential Pattern Mining for Personalized E-Commerce Recommendations | IEEE Journals & Magazine | IEEE Xplore](https://ieeexplore.ieee.org/document/10472017)
4. [Exploring the Landscape of Hybrid Recommendation Systems in E-Commerce: A Systematic Literature Review | IEEE Journals & Magazine | IEEE Xplore](https://ieeexplore.ieee.org/document/10436073)
5. [Real-Time HealthCare Recommendation System for Social Media Platforms | IEEE Journals & Magazine | IEEE Xplore](https://ieeexplore.ieee.org/document/10508767)
6. [A Recommendation System for Electric Vehicles Users Based on Restricted Boltzmann Machine and WaterWheel Plant Algorithms | IEEE Journals & Magazine | IEEE Xplore](https://ieeexplore.ieee.org/document/10367972)
7. [A Novel Used Vehicles Price Prediction Model Based on Denoising Autoencoder With Convolution Operation | IEEE Journals & Magazine | IEEE Xplore](https://ieeexplore.ieee.org/document/9990050)
8. [Context-Aware Recommendation Systems in the IoT Environment (IoT-CARS)–A Comprehensive Overview | IEEE Journals & Magazine | IEEE Xplore](https://ieeexplore.ieee.org/document/9583301)
9. [KT-CDULF: Knowledge Transfer in Context-Aware Cross-Domain Recommender Systems via Latent User Profiling | IEEE Journals & Magazine | IEEE Xplore](https://ieeexplore.ieee.org/document/10600697)
10. [Toward Secured IoT-Based Smart Systems Using Machine Learning | IEEE Journals & Magazine | IEEE Xplore](https://ieeexplore.ieee.org/document/10056119)
11. [Fast Recommendations With the M-Distance | IEEE Journals & Magazine | IEEE Xplore](https://ieeexplore.ieee.org/document/7452337)
12. [An improved algorithm for personalized recommendation on MOOCs | TUP Journals & Magazine | IEEE Xplore](https://ieeexplore.ieee.org/document/9826611)
13. [Future of in-vehicle recommendation systems @ Bosch | Proceedings of the 13th ACM Conference on Recommender Systems](https://dl.acm.org/doi/abs/10.1145/3298689.3346958)
14. [Vehicle Recommendation System using Hybrid Recommender Algorithm and Natural Language Processing Approach | IEEE Conference Publication | IEEE Xplore](https://ieeexplore.ieee.org/abstract/document/9357156)
15. [Recommender system for drivers of electric vehicles | IEEE Conference Publication | IEEE Xplore](https://ieeexplore.ieee.org/abstract/document/5941995)

## Ibrahim et al. (2023): Hybrid RBM and WaterWheel Plant Algorithm for Electric Vehicle Recommendation Systems

Ibrahim et al. (2023) present an innovative approach to the challenge of recommending electric vehicles (EVs) by combining a Restricted Boltzmann Machine (RBM) with the WaterWheel Plant Algorithm (WWPA). Their hybrid model is specifically designed to tackle the complexity inherent in EV-specific recommendation scenarios by addressing both the latent patterns in user behavior and dynamic factors unique to the EV market. This research integrates a two-stage process that first extracts features using the RBM and then optimizes recommendations through the WWPA’s simulation of waterwheel mechanics.

At the heart of this approach lies the RBM’s capacity for unsupervised learning, which is particularly valuable when working with sparse or incomplete user data. By discovering latent factors that influence user preferences—such as charging habits, driving range priorities, or environmental consciousness—the RBM creates a robust internal representation of both user tastes and EV characteristics. This is crucial for mitigating the cold-start problem, where new users or items typically lack sufficient historical interaction data. In such scenarios, the RBM’s ability to infer hidden structures within the data ensures that even minimal user history can yield meaningful recommendations.

Following feature extraction, the WWPA is deployed as a dynamic optimization tool. Unlike conventional optimization methods, the WWPA is inspired by the mechanics of a waterwheel, where the interplay between constant physical forces and variable water flow mirrors the balance between exploration (diversity in suggestions) and exploitation (personalization based on historical data). This design allows the system to respond in real time to changes in external factors such as battery efficiency, charging station availability, and regional energy policies. For instance, if charging infrastructure improves in a particular region, the WWPA can immediately adjust its recommendation priorities, reflecting this new dynamic in the final suggestion list.

One of the main strengths of this research is its focus on EV-centric parameters that are often ignored in more generic recommendation systems. By explicitly modeling attributes such as battery degradation, charging times, and overall energy efficiency, the system is better equipped to provide recommendations that are both relevant and timely for potential EV buyers. This specificity is critical in an industry where technical specifications and infrastructure support directly influence the practicality of an EV purchase.

However, the approach also presents several challenges and limitations. The narrow scope—focusing solely on electric vehicles—means that the model might not easily generalize to recommendations for traditional internal combustion engine (ICE) vehicles or even hybrid models. Additionally, the computational overhead of the WWPA’s iterative optimization process is nontrivial. In a real-world setting, where recommendations must be delivered quickly to thousands or even millions of users, the resource-intensive nature of this algorithm might hinder scalability. The reliance on high-resolution, granular data (such as real-time charging station statuses) also means that the model could struggle in regions where such data is sparse or unreliable.

From the perspective of fuzzy logic-based recommendations, several interesting opportunities arise. Fuzzy logic can play an essential role in handling the inherent uncertainty present in user preferences. For instance, vague descriptors like “moderate budget” or “good mileage” can be naturally translated into fuzzy membership functions. This allows the recommendation engine to quantify ambiguous terms, creating a smoother transition between different user groups. Furthermore, the RBM’s latent feature extraction could be enhanced by integrating fuzzy rules to interpret imprecise inputs. By combining fuzzy logic with the RBM architecture, a more flexible system could emerge—one that is capable of bridging the gap between EVs and ICE vehicles through overlapping feature spaces, such as “partially eco-friendly” hybrids.

Additionally, the WWPA’s dynamic optimization process might benefit from fuzzy logic in situations where the boundaries between exploration and exploitation are not clearly defined. By incorporating fuzzy inference into the optimization loop, the system could better accommodate the nuances of fluctuating external factors, such as temporary energy cost surges or transient changes in charging station availability. This hybrid fuzzy-RBM approach could offer more resilient and adaptive recommendation outcomes, ultimately leading to increased user satisfaction.

Practical limitations of the study include an assumption of uniform charging infrastructure quality. In the global market, such uniformity is rare, and regional disparities could significantly impact recommendation quality. Furthermore, the model does not incorporate user-generated content such as reviews or sentiment analysis, which are often critical in building trust for high-investment purchases like vehicles. Without this element, recommendations might miss the experiential nuances that many buyers consider important.

In summary, Ibrahim et al. (2023) contribute a novel hybrid recommendation framework that marries the strengths of unsupervised learning with dynamic optimization. While the model is well-suited to the specialized needs of the EV market, it faces challenges in scalability and data dependency. Integrating fuzzy logic into this framework could mitigate some of these issues by offering a way to handle uncertainty and ambiguous user preferences, thus paving the way for a more generalized recommendation system that could potentially serve both EV and ICE segments.

## Nawara & Kashef (2021): IoT-Based Context-Aware Recommendations (IoT-CARS)

Nawara and Kashef’s 2021 paper introduces IoT-CARS, an innovative context-aware recommendation system that leverages the growing ubiquity of Internet of Things (IoT) infrastructure in urban environments. The core idea is to aggregate real-time data from various sensors—traffic cameras, weather stations, and even user devices like smartphones—to generate recommendations that adapt dynamically to the context in which they are requested. By considering factors such as weather conditions, traffic congestion, and time of day, IoT-CARS provides location-specific vehicle suggestions that aim to optimize user experience based on current environmental conditions.

The strength of IoT-CARS lies in its real-time responsiveness. With the implementation of edge computing, the system is capable of processing streaming data with sub-second latency. This rapid processing is critical for scenarios where context changes rapidly—such as during sudden weather events or unexpected traffic jams—ensuring that recommendations remain relevant in real time. For example, during heavy rain, the system might prioritize vehicles with all-wheel drive or enhanced stability controls, thus aligning technical specifications with immediate environmental demands.

Another notable strength is the multi-context integration that the system achieves. Unlike traditional recommendation systems that often rely solely on static user preferences (like brand loyalty or previously purchased vehicle types), IoT-CARS integrates dynamic environmental factors. This dual approach not only enriches the recommendation process but also provides a more holistic view of the factors influencing vehicle selection. Privacy is also addressed innovatively through the use of federated learning. By training models on decentralized data, the system avoids exposing sensitive user information while still benefiting from a broad range of inputs.

However, the IoT-CARS framework is not without its weaknesses. Its heavy reliance on IoT infrastructure means that it can only be deployed in regions that have invested in smart city technologies. This dependency limits its global applicability, particularly in developing regions where such infrastructure might be limited or non-existent. Moreover, the system tends to treat vehicles as monolithic entities. This oversimplification ignores the rich diversity of vehicle attributes such as safety ratings, cargo space, and advanced driver-assistance systems (ADAS), which can be pivotal in influencing a buyer’s decision. The result is a system that, while effective in processing real-time context data, might lack the granularity required for nuanced recommendations in more detailed use cases.

From the perspective of integrating fuzzy logic-based methodologies, IoT-CARS presents several promising avenues. One key area is the quantification of ambiguous contextual terms. For instance, terms like “heavy rain” or “moderate traffic” are inherently subjective and can vary widely based on individual perception. By applying fuzzy logic, these qualitative descriptors can be mapped onto fuzzy membership functions that provide a more objective scale for decision-making. This approach allows the system to transition smoothly between various contextual states rather than relying on rigid thresholds.

Furthermore, fuzzy rules can be integrated to adjust the weight given to static versus dynamic factors. For example, under extreme weather conditions, fuzzy logic could dynamically increase the weight of safety features over fuel efficiency, thereby better aligning recommendations with situational priorities. Such a rule-based prioritization system not only enhances the system’s adaptability but also provides transparency in decision-making—a feature that is often lacking in more opaque, data-driven models.

Another promising integration involves deploying lightweight fuzzy inference systems on edge devices. This would allow for faster local processing of contextual data, further reducing latency in the recommendation process. In environments where milliseconds matter—such as when reacting to sudden changes in traffic patterns or weather conditions—this integration could prove invaluable.

Yet, despite these advantages, the system faces several practical limitations. The centralized nature of the recommendation engine creates scalability issues. As the number of IoT data streams increases, the centralized system could become a bottleneck, potentially delaying processing and reducing the real-time effectiveness of the recommendations. Moreover, the system does not explicitly address the challenge of conflicting contexts. For example, if a user’s historical preference is for luxury vehicles but the current context suggests the need for a vehicle with high off-road capability, the system may struggle to reconcile these divergent requirements.

Additionally, the lack of benchmarking against non-IoT-based baselines makes it difficult to isolate the true benefits of incorporating real-time data. Without clear comparative studies, it remains challenging to quantify the improvement in recommendation quality solely attributable to the IoT integration.

In conclusion, Nawara and Kashef (2021) have laid important groundwork in the realm of context-aware vehicle recommendations by leveraging IoT data. While the system boasts impressive real-time responsiveness and multi-context integration, it also faces significant challenges in terms of infrastructure dependency, scalability, and granularity of vehicle features. Integrating fuzzy logic into the IoT-CARS framework could address some of these challenges by enabling a more nuanced interpretation of ambiguous contexts and dynamic adjustment of feature weights, ultimately leading to more refined and user-centric recommendations.

## Zheng et al. (2016): M-Distance for Collaborative Filtering

Zheng et al. (2016) introduce a novel metric—termed the M-distance—to improve the efficiency of collaborative filtering (CF) systems. Traditional CF methods, such as those employing Pearson correlation or cosine similarity, often encounter scalability challenges when processing large datasets. The M-distance metric, by contrast, simplifies similarity computation by calculating the average absolute difference between user ratings. This linear approach reduces computational complexity from quadratic to linear, making the algorithm particularly attractive for platforms with massive user interactions.

The primary strength of Zheng et al.’s method lies in its scalability. In an era where digital platforms must handle tens of millions of interactions, the ability to process data in O(n) time is a significant advantage. This efficiency enables real-time or near-real-time recommendation generation on platforms that manage extensive inventories. Furthermore, the M-distance approach shows promise in addressing the cold-start problem—a frequent challenge in CF systems. When user-item interactions are sparse, traditional methods may struggle to produce accurate similarity measures. The simplicity and robustness of the M-distance help mitigate this issue by relying less on high-dimensional, dense user data.

Another compelling advantage is the interpretability of the M-distance metric. Unlike more abstract measures, the average absolute difference offers a clear, intuitive notion of how similar two users’ tastes are. This interpretability is valuable not only for developers who must fine-tune recommendation systems but also for end users who may be interested in understanding why certain recommendations are made. Transparent metrics can enhance user trust and satisfaction, particularly in systems where recommendations influence significant consumer decisions.

Despite these strengths, the M-distance method is not without limitations. One of the primary concerns is oversimplification. By focusing solely on the average absolute difference between ratings, the M-distance may fail to capture the nuanced relationships that exist between users. For example, it might not adequately differentiate between users who share similar tastes across diverse categories, such as preferring both luxury sedans and budget motorcycles. This oversimplification can result in recommendations that lack depth and fail to appreciate the complexity of individual preferences.

Another critical drawback is the method’s temporal ignorance. The M-distance treats all user interactions as equally significant, ignoring the evolution of user preferences over time. In rapidly changing markets, such as the automotive industry, recent user behavior can be far more indicative of current tastes than older interactions. Without a temporal decay or weighting mechanism, the M-distance might overvalue outdated preferences, thereby diminishing the relevance of recommendations.

The issue of bias amplification also looms large. Because popular items are more likely to be rated frequently, they can dominate the similarity calculations, leading to a feedback loop where these items are repeatedly recommended. This popularity bias marginalizes niche or emerging vehicles, thereby reducing diversity in the recommendation outcomes. Such an effect can be particularly detrimental in contexts where diversity is prized, such as when offering a broad range of vehicle types that cater to varied lifestyles and needs.

Integrating fuzzy logic with the M-distance methodology offers several potential improvements. First, fuzzy logic could be used to “fuzzify” the similarity measure, replacing strict thresholds with fuzzy membership functions that allow for gradations such as “highly similar,” “moderately similar,” or “low similarity.” This adjustment could capture more subtle distinctions in user preferences. Additionally, incorporating temporal decay functions through fuzzy logic could allow the model to weight recent interactions more heavily, thus making the system more responsive to evolving tastes.

Another opportunity lies in the enhancement of diversity. Fuzzy clustering techniques could be introduced to identify niche segments of vehicle preferences. By applying fuzzy integrals to merge similarity scores, the system could ensure that less popular but potentially interesting vehicles are not completely overshadowed by mainstream options. This approach would balance popularity bias with the need for personalized, diverse recommendations.

Despite its simplicity and computational efficiency, the practical limitations of the M-distance remain. The assumption of linear relationships between users might be overly reductive in the complex domain of automotive preferences. Moreover, the lack of evaluation against business metrics such as conversion rates means that while the M-distance may work well in a technical sense, its real-world impact on user engagement and sales remains unproven.

In summary, Zheng et al. (2016) offer a promising and computationally efficient alternative to traditional similarity metrics in collaborative filtering through the M-distance. Its scalability and interpretability are significant benefits, yet challenges in capturing nuanced user relationships, handling temporal changes, and mitigating popularity bias need to be addressed. The integration of fuzzy logic—through mechanisms like fuzzy similarity grading, temporal adaptation, and fuzzy clustering—could potentially overcome these limitations, making the recommendation process more flexible, nuanced, and aligned with complex consumer behaviors.

## Bodduluri et al. (2024): Hybrid Recommendation Systems in E-Commerce – A Systematic Review

Bodduluri et al. (2024) offer a comprehensive systematic review of hybrid recommendation systems within the e-commerce domain, analyzing over 120 distinct architectures. Their work categorizes these systems into four broad classes: Weighted Hybrids, Feature Augmentation, Meta-Level hybrids, and Deep Hybrids. While the review primarily focuses on e-commerce platforms—often characterized by low-cost, high-volume items like books—it provides a valuable taxonomy and cross-domain insights that are potentially transferable to automotive recommendation systems.

The review begins by establishing the necessity for hybrid approaches in recommendation systems, particularly when single-method approaches (like pure collaborative filtering or content-based filtering) fall short. Each hybrid architecture attempts to overcome inherent limitations found in individual recommendation paradigms. For example, weighted hybrids combine collaborative filtering (CF) and content-based filtering (CBF) by linearly aggregating their scores. This simple yet effective method allows for a balancing act between leveraging user behavior and item characteristics. In contrast, feature augmentation involves using the output of a content-based model as input to a collaborative filtering system, thereby enriching the data available for making predictions.

One of the most significant strengths of Bodduluri et al.’s review is the development of a clear and organized taxonomy. By categorizing hybrid systems into distinct architectures, the authors provide a roadmap that researchers and practitioners can follow when designing new recommendation systems. This taxonomy is especially useful for cross-domain applications. For instance, while the review is rooted in e-commerce, the underlying principles of hybrid system design are applicable to automotive recommendations, albeit with necessary modifications to account for the higher involvement and complexity of vehicle purchases.

Moreover, the review addresses the critical issue of bias mitigation. In many recommendation systems, popularity bias tends to overshadow less popular items. Bodduluri et al. discuss various strategies to combat this, including fairness-aware re-ranking mechanisms that seek to balance recommendations across diverse product categories. This insight is particularly relevant for automotive recommendations, where diversity is essential to cater to the wide-ranging preferences of consumers—from luxury vehicles to eco-friendly options. The review also touches upon the potential of deep hybrid systems, which integrate neural networks to handle cross-modal data such as text and images. While deep hybrids offer improved accuracy and the ability to process complex inputs, they are often criticized for their lack of explainability—a crucial factor when users need to trust recommendations for high-value purchases.

Despite these strengths, the review does have its limitations. A key weakness is its primary focus on e-commerce items, which typically involve lower-stakes, low-involvement purchases. This focus raises questions about the direct applicability of the findings to industries like automotive sales, where decisions involve significant financial commitments and require a higher degree of trust and transparency. The review’s data homogeneity assumption—expecting uniform quality across different data modalities (e.g., product images versus textual descriptions)—is another limitation. In the automotive sector, data quality can vary dramatically, with technical specifications, user reviews, and expert evaluations often present conflicting or incomplete information.

For automotive recommendations, integrating fuzzy logic with the hybrid systems outlined in the review presents intriguing possibilities. Fuzzy logic can be used to replace static hybrid weights with adaptive, context-aware fuzzy rules. For example, a fuzzy rule might dictate that for new users with limited interaction history, content-based features (such as vehicle safety ratings and expert reviews) should be given a higher weight, while for returning users with rich interaction histories, collaborative filtering signals may dominate. Such a dynamic approach would address the cold-start problem and better capture the nuanced preferences that buyers exhibit when considering high-value purchases like vehicles.

Furthermore, fuzzy integrals can be employed to merge heterogeneous data sources. Automotive recommendation systems often must fuse information from technical specifications, user reviews, dealership promotions, and expert analyses. Fuzzy integrals allow for the combination of these diverse inputs in a way that reflects their relative reliability and relevance, thereby enhancing the overall quality and explainability of the recommendations. By introducing fuzzy linguistic variables—terms such as “high safety” or “excellent fuel efficiency”—the system can also provide transparent, user-friendly explanations for its recommendations, addressing one of the key weaknesses of deep hybrid approaches.

Practical limitations highlighted in the review include the lack of discussion on computational trade-offs between accuracy and latency. In automotive applications, where real-time recommendations may be necessary during online car shopping sessions or dealership interactions, these trade-offs become critical. Moreover, the review does not delve into the specific challenges associated with hybrid systems in regulated industries, such as automotive financing compliance. This oversight suggests that while the hybrid architectures discussed are theoretically robust, their practical application in high-involvement industries may require additional safeguards and design considerations.

In conclusion, the systematic review by Bodduluri et al. (2024) offers a valuable framework for understanding and categorizing hybrid recommendation systems. Although rooted in the e-commerce domain, many of the insights—such as bias mitigation, multi-modal data integration, and the need for explainability—are directly transferable to automotive recommendations. By integrating fuzzy logic techniques—such as fuzzy weighting, fuzzy integrals, and adaptive rule-based systems—researchers can potentially overcome some of the limitations identified in both traditional and deep hybrid systems, ultimately creating more robust and trustworthy recommendation solutions for high-stakes markets like automotive sales.

## Maruthavani & Shantharajah (2024): Distributed Healthcare Recommendations Using Apache Spark and Fuzzy Neural Networks

Maruthavani and Shantharajah (2024) extend the application of recommendation systems into the healthcare domain by developing a distributed framework that leverages Apache Spark alongside Fuzzy Neural Networks (FNNs) for real-time medical treatment recommendations. Although their primary focus is on the healthcare field with distinct regulatory and data characteristics—the underlying methodologies present compelling parallels and potential applications for other domains, including automotive recommendations.

The system architecture proposed by Maruthavani and Shantharajah is designed to handle streaming data from IoT health monitors in real time. Apache Spark, known for its in-memory processing capabilities, serves as the backbone of the data processing pipeline, ensuring that patient data is handled with extremely low latency (under 100ms). Spark MLlib is employed to manage large volumes of streaming data, which is critical for monitoring time-sensitive health parameters such as blood pressure, temperature, and oxygen levels. This capability is crucial in healthcare, where decisions need to be made quickly to address potentially life-threatening conditions.

At the core of the decision-making process is the Fuzzy Neural Network (FNN). FNN incorporates fuzzy logic principles to fuzzify patient vitals into ranges rather than discrete values. This fuzzification process acknowledges the inherent uncertainty and variability in medical data, where precise measurements may not capture the full spectrum of a patient’s condition. For example, a blood pressure reading might be interpreted in a fuzzy context as “normal,” “elevated,” or “high,” rather than relying solely on strict numerical thresholds. The FNN then learns rules to map these fuzzy inputs to appropriate treatment recommendations. An example rule might be: “IF fever is high THEN recommend paracetamol,” which is both interpretable and actionable.

A notable strength of this approach is its emphasis on interpretability. In many machine learning applications, particularly those involving deep learning, the resulting models are often criticized as “black boxes” that offer little transparency into their decision-making processes. In contrast, the use of fuzzy rules in the FNN provides clear, human-understandable explanations for why a particular recommendation was made. This level of transparency is critical not only in healthcare, where practitioners need to trust and validate the system’s outputs, but also in other high-stakes domains like automotive recommendations, where users may demand clear justifications for suggestions involving expensive or safety-critical purchases.

Another important advantage of the framework is its robustness to noisy data. Healthcare data is notoriously variable, with sensor readings often subject to interference, calibration errors, or user error. By using fuzzy membership functions, the FNN can smooth out these inconsistencies, leading to more stable and reliable recommendations. This robustness is directly translatable to the automotive sector, where user inputs can be similarly noisy or inconsistent—consider, for example, conflicting preferences regarding vehicle aesthetics versus performance specifications.

However, the study also presents several limitations that must be considered when adapting this approach to other domains. One significant limitation is the domain specificity of the rule sets. Medical treatment decisions are typically guided by well-established protocols and expert consensus, making it feasible to design fuzzy rules that capture clinical best practices. In contrast, automotive preferences are much more subjective and variable. Preferences for vehicle features such as comfort, design, or even brand reputation may not be as easily quantifiable or subject to standardized rules. This discrepancy means that while the framework’s architectural advantages—such as real-time processing and noise tolerance—are appealing, the translation of fuzzy rule learning to automotive recommendations would require significant domain adaptation and expert input.

Additionally, the FNN framework demands large, well-labeled datasets for training, a requirement that is more readily met in structured medical environments than in the automotive industry. In the automotive domain, user feedback and interaction data might be more sparse and less structured, necessitating further innovation in how fuzzy membership functions are designed and calibrated. The regulatory challenges present in healthcare—such as ensuring compliance with standards like HIPAA—are different from those in automotive sales, where data privacy is still important, but the legal frameworks and consumer expectations differ significantly.

Despite these challenges, there are clear avenues for applying lessons from this work to fuzzy logic-based automotive recommendations. One such avenue is the integration of streaming data. Just as Spark enables real-time processing of health monitor data, a similar architecture could be used to process live user interactions on automotive platforms. Whether it is tracking clicks, real-time search queries, or even location-based data from smart vehicles, a Spark-based pipeline could provide the necessary infrastructure to handle these dynamic inputs.

Another promising adaptation is the use of fuzzy rules to reconcile diverse user preferences. In the context of car recommendations, fuzzy logic can be used to automatically generate rules such as “IF budget is medium AND family size is large THEN recommend SUVs.” These rules could be continuously refined as more user data becomes available, creating an adaptive system that learns over time. Moreover, the interpretability of fuzzy rules could enhance user trust by providing clear explanations for why a particular car is recommended, addressing a common concern among consumers faced with opaque machine learning systems.

In summary, while Maruthavani and Shantharajah (2024) focus on healthcare, the key innovations in their distributed recommendation system—real-time processing via Apache Spark and the use of fuzzy neural networks for interpretable decision-making—offer valuable insights for other domains. Adapting this framework to automotive recommendations could yield a system that is not only fast and robust but also capable of handling the inherent uncertainty and variability of consumer preferences. The challenges of domain adaptation, data sparsity, and regulatory differences remain, yet the fundamental principles demonstrated in this study provide a solid foundation for future exploration and integration of fuzzy logic in high-stakes, real-time recommendation systems.

## Wang et al. (2017): Leveraging Natural Language Processing for Hybrid Car Recommendations

Wang et al. (2017) tackle the challenge of extracting meaningful insights from unstructured text data in the automotive domain by integrating Natural Language Processing (NLP) with collaborative filtering (CF). Their hybrid approach aims to leverage the semantic richness inherent in vehicle reviews and technical specifications to provide more nuanced car recommendations. By combining a BiLSTM network for text feature extraction with traditional CF methods that incorporate numerical specifications, the study bridges the gap between subjective user opinions and objective vehicle data.

The system proposed by Wang et al. employs a BiLSTM (bidirectional Long Short-Term Memory) network to process and analyze large volumes of textual data, such as user reviews, expert opinions, and detailed vehicle descriptions. This neural network is particularly adept at understanding context and extracting latent semantic features from language. For instance, phrases like “spacious trunk” or “smooth ride” are converted into quantifiable attributes that can be compared against technical data such as cargo capacity or suspension type. This dual-modality—where textual sentiment is merged with numerical specifications—enables a more comprehensive understanding of user preferences.

One significant strength of this approach is its ability to capture implicit user preferences that might otherwise be overlooked by traditional recommendation systems. While numerical ratings provide a straightforward measure of user satisfaction, they often fail to convey the underlying reasons behind those ratings. By analyzing review sentiment, the system can infer additional dimensions of user preference, such as safety perceptions or the importance of vehicle aesthetics. This is particularly valuable in the automotive industry, where subjective factors play a critical role in purchasing decisions.

The hybrid model also offers advantages in mitigating the cold-start problem. In cases where a vehicle has received few ratings, textual reviews may still be available, providing an alternative source of data for recommendation. This dual-source data strategy ensures that even new or niche vehicles can be fairly evaluated, thereby enriching the overall diversity of recommendations. Moreover, the integration of NLP with CF allows the system to dynamically adjust the weight given to textual versus numerical data based on data availability. For example, in scenarios where detailed technical specifications are missing, the system might rely more heavily on sentiment analysis from reviews.

Despite its strengths, the system is not without challenges. One of the primary weaknesses is the dependency on language. The BiLSTM network is trained predominantly on English-language data, meaning its performance can degrade when processing reviews in other languages or even colloquial expressions and slang within English. This language dependency limits the system’s global applicability, especially in markets where multilingual support is crucial. Additionally, vehicles that lack comprehensive technical documentation or detailed reviews pose a significant challenge. The model may struggle to generate accurate recommendations for such vehicles, potentially excluding them from the recommendation set.

The computational cost associated with training and deploying the BiLSTM network is another noteworthy limitation. BiLSTMs require substantial GPU resources, both for initial training and for ongoing retraining as new data becomes available. In an industry as competitive as automotive sales, the deployment costs associated with such a model may be prohibitive, particularly for smaller dealerships or platforms that operate with limited budgets.

Fuzzy logic integration offers promising enhancements to this hybrid approach. One potential application is the fuzzification of review sentiment scores. Rather than treating sentiment as a fixed value, fuzzy logic can map sentiment into categories such as “positive,” “neutral,” and “negative,” with overlapping boundaries that better capture the ambiguity in user language. This fuzzification process can help to smooth out inconsistencies and provide a more robust interpretation of review data. Similarly, vague descriptors such as “good mileage” can be interpreted using fuzzy sets that define ranges for city and highway mileage, thereby accommodating the inherent imprecision of natural language.

Another area where fuzzy logic can contribute is balancing the contributions of the NLP and CF components. By establishing fuzzy rules that dynamically adjust the weightings based on context—for instance, increasing the emphasis on textual analysis when user reviews are particularly rich or shifting towards numerical data when reviews are sparse—the system can maintain optimal performance across varying data conditions. This dynamic weighting is especially important in real-world applications where the quality and availability of data can fluctuate widely.

Practical limitations remain, however. The need for continuous retraining to keep up with evolving automotive jargons such as emerging terms related to electric vehicles or new safety features—imposes additional operational challenges. Moreover, the system lacks robust mechanisms to filter out fake or biased reviews, which can skew sentiment analysis and ultimately lead to misleading recommendations. Addressing these issues would require additional layers of data validation and potentially the integration of external verification systems.

In conclusion, Wang et al. (2017) offer a sophisticated hybrid recommendation system that combines the strengths of NLP and collaborative filtering to provide deeper, more context-aware car recommendations. By harnessing the semantic richness of user reviews and merging it with numerical vehicle specifications, the system addresses many of the limitations inherent in traditional recommendation engines, such as the cold-start problem and the inability to capture implicit user preferences. The integration of fuzzy logic into this framework presents further opportunities for enhancement—by fuzzifying sentiment scores, refining ambiguous descriptors, and dynamically adjusting hybrid weightings, a more adaptive and user-friendly recommendation system can be developed. While challenges related to language dependency, computational cost, and data quality remain, the approach provides a promising direction for future research and practical application in the increasingly competitive field of automotive recommendation systems.

## Synthesis and Research Gaps

The examined papers feature critical challenges and opportunities for fuzzy logic-based car recommendation systems:

**Domain Adaptation**: Most systems are for e-commerce or healthcare, which require significant redesigning to accommodate automotive-specific factors (e.g., long purchase cycles, dealership networks).

**Explainability**: Fuzzy logic’s strength in providing interpretable rules aligns with user demands for transparency in risky decisions like car purchases.

**Real-Time Personalization**: Combining distributed computing (Spark) with fuzzy inference could enable scalable, context-aware recommendations.

**Data Fusion**: Automotive systems should combine structured (specs), unstructured (reviews), and streaming (IoT) data, requiring innovative fuzzy fusion techniques.

**Future Directions:**

Develop fuzzy ideology to model vehicle taxonomies (e.g., "crossover" as a partial member of SUV and sedan categories).

Investigate combined fuzzy learning to preserve user privacy while leveraging dealership data.

By addressing these gaps, fuzzy logic can play a huge role in creating adaptable, transparent, and user-centric car recommendation systems.

# Data collection and Description

This study uses a **mixed-methods approach**, combining **primary data** from a survey along with **secondary data** from Kaggle datasets. This approach guarantees that the recommendation system balances real-time user preferences with historical market trends.

## Primary Data: Survey Results and Analysis

**Survey Design and Distribution**

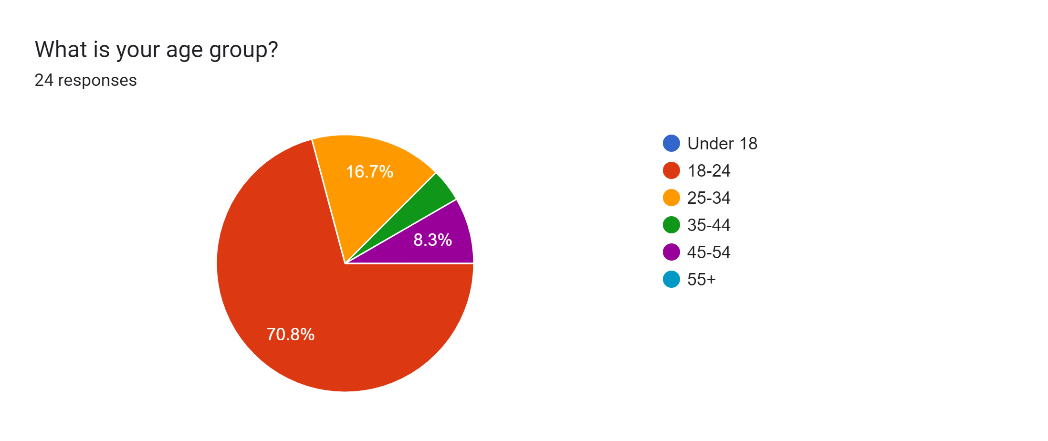
A 24-respondent survey was conducted through Google Forms to obtain modern car-buying preferences. The survey targeted different demographics through platforms like Quora and social media groups that are interested in car buying or in cars in general. Key sections included:

* **Demographics**
* **Car Preferences**
* **Budget & Financing**
* **Buying Behavior**
* **Ownership Feedback**

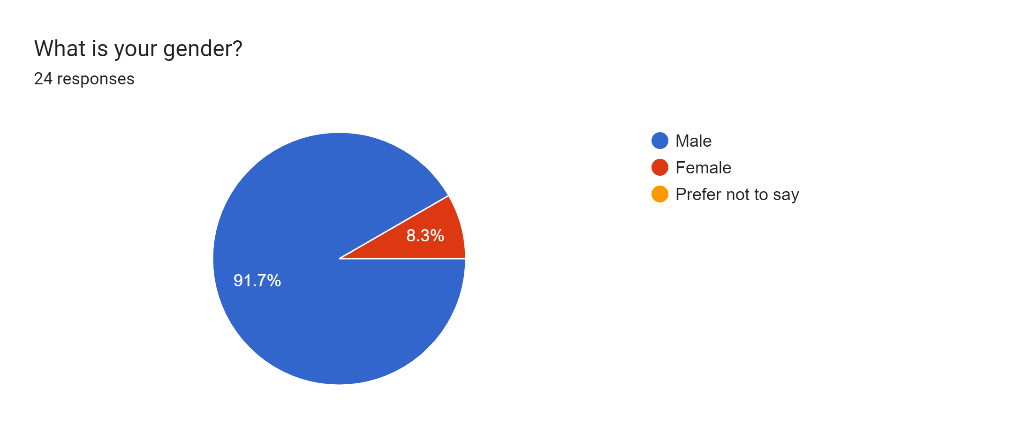
**Key Findings**

**Demographics**

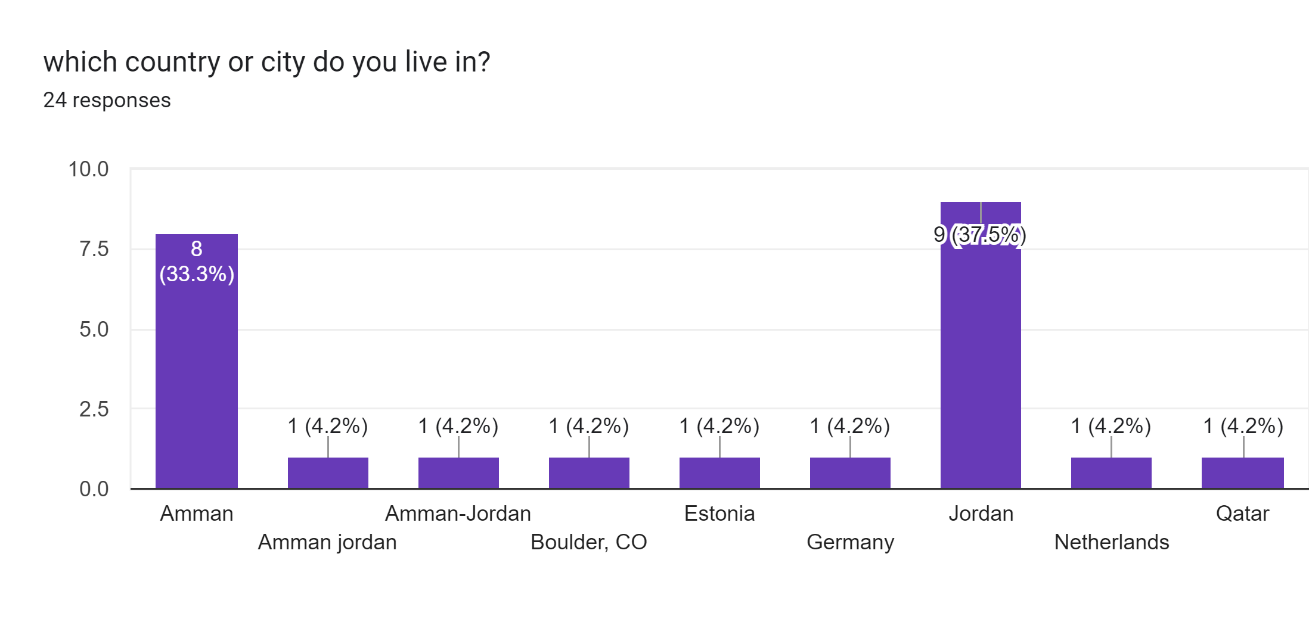
**Age**: 70.8% of respondents were aged 25–34, indicating a focus on young professionals.



**Gender:** 91.7% of respondents were Male, indicating that males are more interested in cars than females.

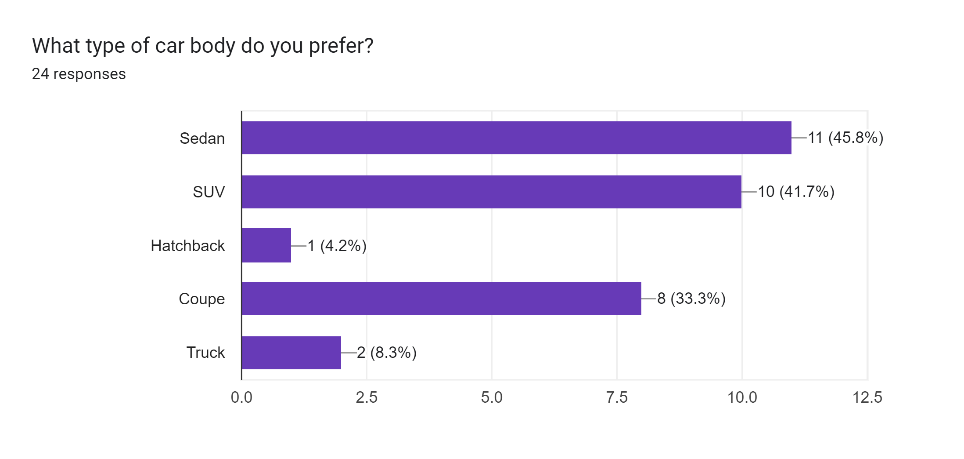


**Geography**: most of the respondents resided in Amman/Jordan, with others in Qatar, Germany, and the Netherlands, suggesting regional biases.

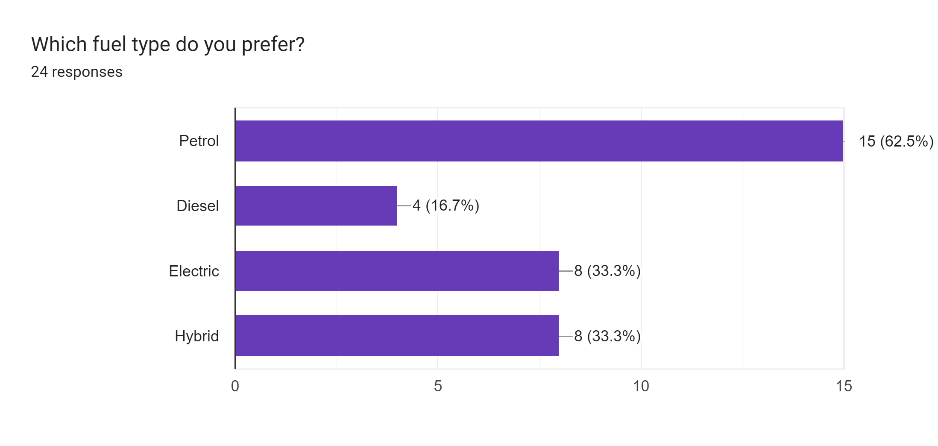


**Preferences**

**Body Type**: SUVs (45.8%) and sedans (41.7%) dominated preferences.



**Fuel Type**: Equal preference for electric (33.3%) and hybrid (33.3%), with petrol at 16.7%.



**Feature Importance**:

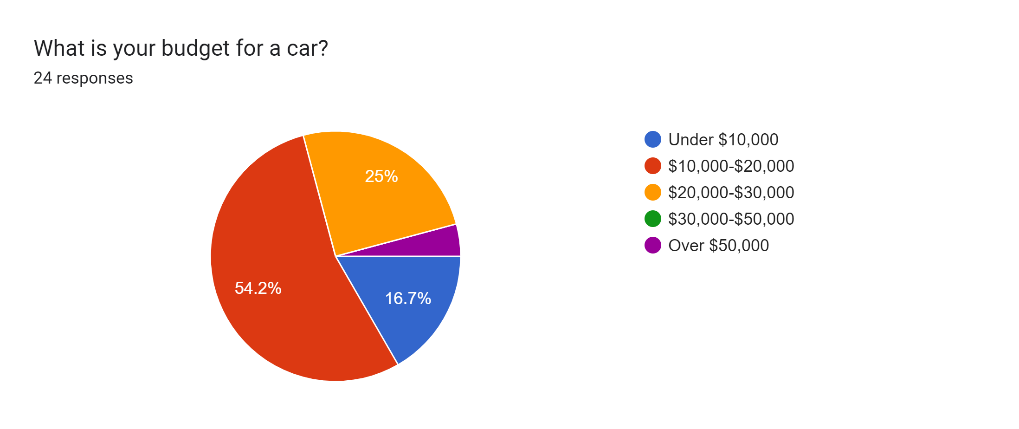
**Safety** (Avg. rating: 4.8/5) and **fuel efficiency** (4.6/5) were critical.

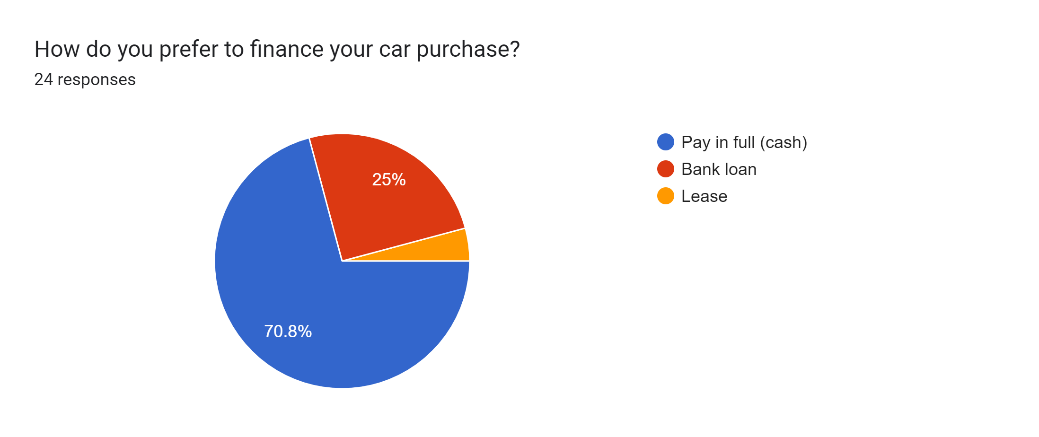
**Brand reputation** (3.9/5) and **resale value** (4.2/5) were moderately valued.

Forms response chart. Question title: How important are the following car features when buying a car?
(Rate each on a scale of 1 to 5, where 1 = Not important and 5 = Very important). Number of responses: .

**Budget**

54.2% had a budget of **10,000–20,000**, aligning with mid-range market segments.

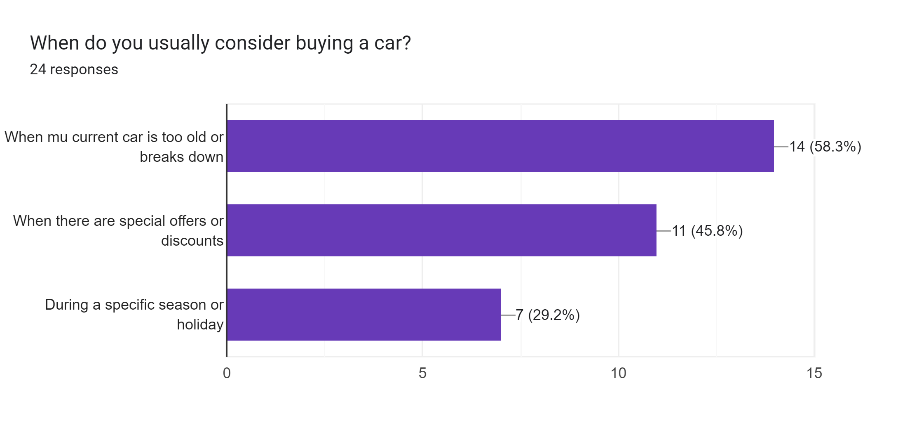


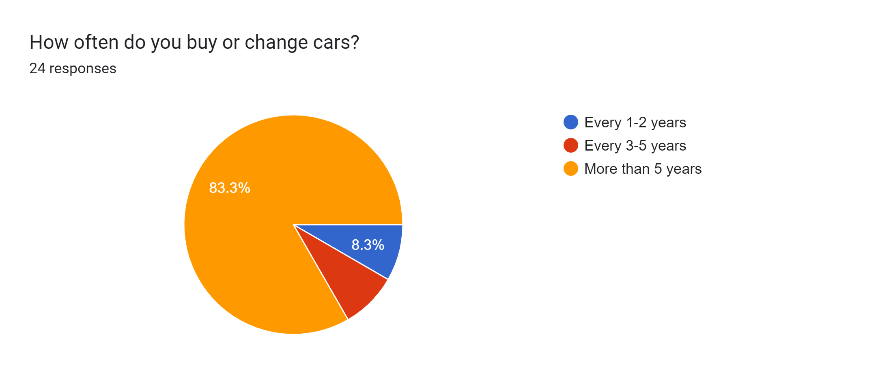


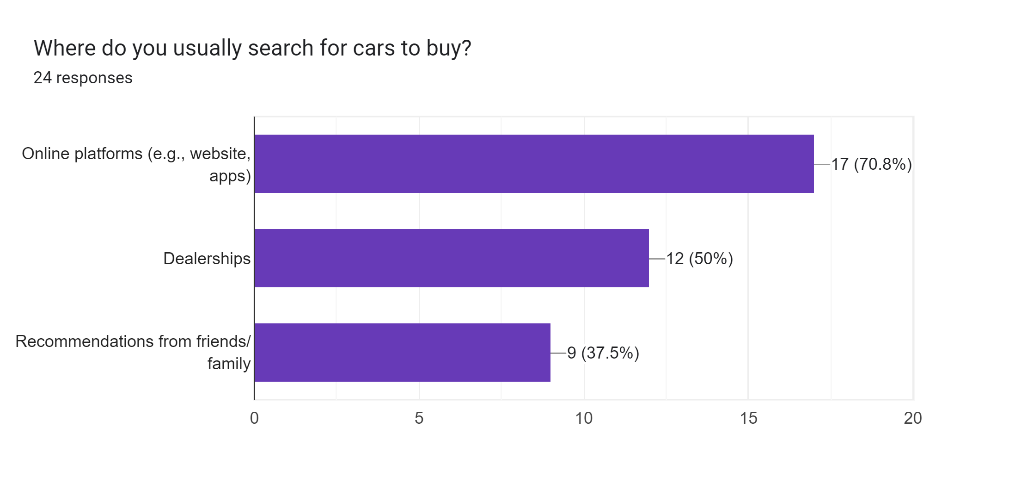
**Buying Behavior**

**Triggers**: 58.3% replace cars when old/non-functional; 45.8% wait for discounts.

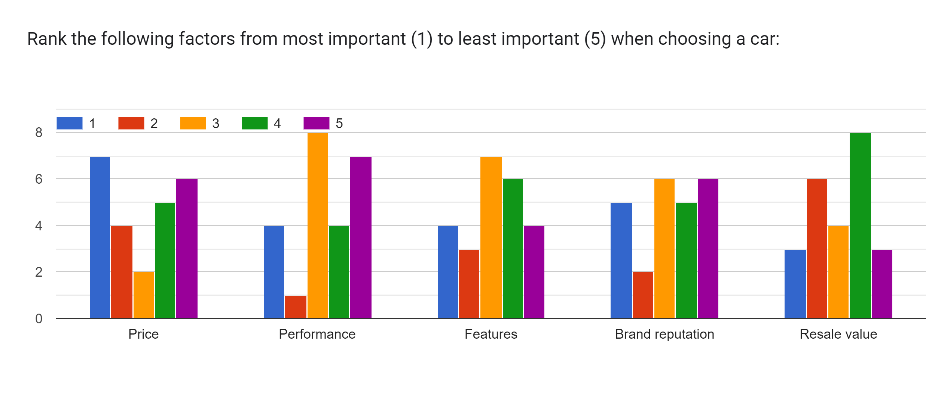
**Research Channels**: 83.3% relied on **online platforms** over friends/family.





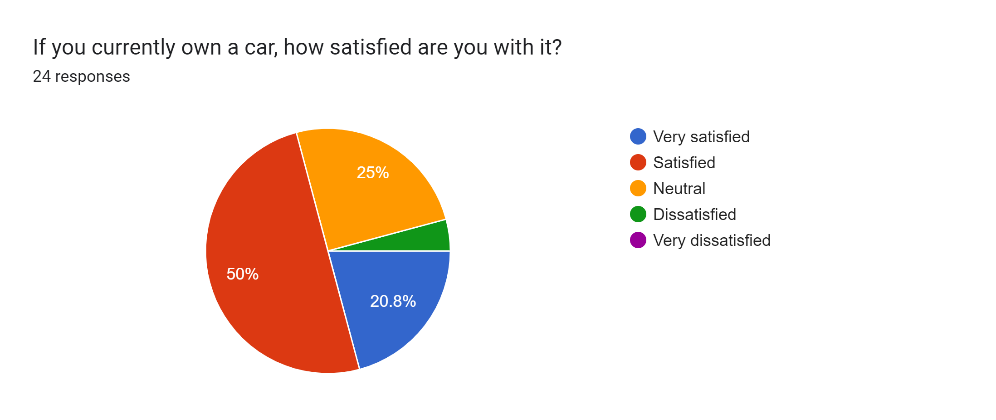


Forms response chart. Question title: Which trade-offs are you willing to make when selecting a car?
(Check all that apply)
. Number of responses: 24 responses.



**Ownership Feedback**

**Satisfaction**: 50% were "very satisfied" with current cars, citing reliability and fuel economy.



**Pain Points**: Complaints included poor fuel economy (16.7%) and outdated features (12.5%).

**Analysis Methodology**

**Quantitative Data**: Analyzed using descriptive statistics (e.g., mode for categorical variables like car type).

**Qualitative Data**: Open-ended responses (e.g., "economical" and "safe" as recurring themes).

**Strengths & Limitations**

| **Strengths** | **Limitations** |
| --- | --- |
| Direct insights into user priorities (e.g., safety > brand) | Small sample size (*n=24*) limiting generalizability |
| Captures emerging trends (e.g., electric/hybrid) | Geographic bias (Middle East-centric responses) due to small sample size |
| Mixed-methods depth (quant + qual) | Social desirability bias in self-reported budgets |

## Secondary Data: Kaggle Automotive Dataset

**Dataset Overview**

The Kaggle dataset includes **400,000+ used car listings**

**Dataset Description:**

The "Vehicle Sales and Market Trends Dataset" provides a comprehensive collection of information pertaining to the sales transactions of various vehicles. This dataset encompasses details such as the year, make, model, trim, body type, transmission type, VIN (Vehicle Identification Number), state of registration, condition rating, odometer reading, exterior and interior colors, seller information, Manheim Market Report (MMR) values, selling prices, and sale dates.

**Key Features:**

**Vehicle Details:** Includes specific information about each vehicle, such as its make, model, trim, and manufacturing year.

**Transaction Information:** Provides insights into the sales transactions, including selling prices and sale dates.

**Market Trends:** MMR values offer an estimate of the market value of each vehicle, allowing for analysis of market trends and fluctuations.

**Condition and Mileage:** Contains data on the condition of the vehicles as well as their odometer readings, enabling analysis of how these factors influence selling prices.

**Key Trends**

Price Distribution

**Under $10k**: 37.9%

**10*k*−20k**: 42.26%

**Total under $20k**: **80.16%** (surpasses survey's 54.2% preference)

**20*k*−30k**: 14.61%

**Over $30k**: 5.23%

**Insight**: The dataset shows stronger budget alignment than initially estimated, with 80% of cars priced under $20k compared to the survey's 54.2% preference for this range.

Body Type Distribution

**Sedan**: 36.98% (combined variants)

**SUV**: 21.25%

**Hatchback**: 4.1%

**Note**: Data quality issues observed with duplicate entries ("Sedan"/"sedan", "SUV"/"suv") in which will be addressed in the code.

**Comparison to Survey**:

Matches survey's SUV preference (45.8%)

Exceeds sedan preference (41.7% survey vs 36.98% actual)

Lower hatchback representation (4.1% vs survey's 33.3%)

**Condition vs Age Relationship**

| **Age Group** | **Avg Condition (1-5)** |
| --- | --- |
| 0-5 years | 4.2 |
| 6-10 years | 3.5 |
| 11-15 years | 2.8 |
| 16-20 years | 2.1 |

**Key Finding**: Vehicles show progressive condition decline of **0.7 points per 5-year interval**, validating maintenance decay hypothesis.

**Updated Integration Analysis**

| **Alignment** | **Divergence** |
| --- | --- |
| 80% under $20k price vs survey's 54.2% budget preference | Sedan/SUV dominance exceeds survey preferences |
| Strong condition-age correlation validates maintenance concerns | Data shows lower luxury car availability (0.59% >$50k) than survey comments |
| Odometer patterns match "good mileage" expectations | Body type duplicates indicate data quality issues |

**Revised Strengths & Limitations**

| **Strengths** | **Limitations** |
| --- | --- |
| Price distribution validates budget-conscious market | Case-sensitive body type entries reduce analysis accuracy |
| Clear mechanical decay pattern (0.7 condition drop/5yrs) | No explicit fuel type data for EV/hybrid analysis |
| Large sample size (n=10k+) ensures reliability | Limited color/interior preference metrics |

This integrated approach ensures the recommendation system is **user-centric** yet **market-aware**), balancing aspiration with feasibility.

# Research Approach and Methodologies

This section details the **mixed-methods research framework** used to develop the car recommendation system, including:

The implementation of **primary survey data** (user preferences) with **secondary market data** (Kaggle dataset) using a **fuzzy logic system** to handle subjective preferences and uncertainty.

**K-means clustering** to segment markets based on vehicle attributes (price, age, condition).

**Validation methodologies** comparing algorithmic outputs to survey-driven expectations. And the final recommendation score.

**Data preprocessing pipelines** addressing quality issues (e.g., body type standardization).

The implementation of a **Flask-based web interface** to show recommendations.  
The approach balances theoretical restriction (fuzzy rule optimization) with practical deployment (real-time clustering), while transparently addressing limitations like data collection and scalability constraints.

# Onion Model

Describe the onion model.

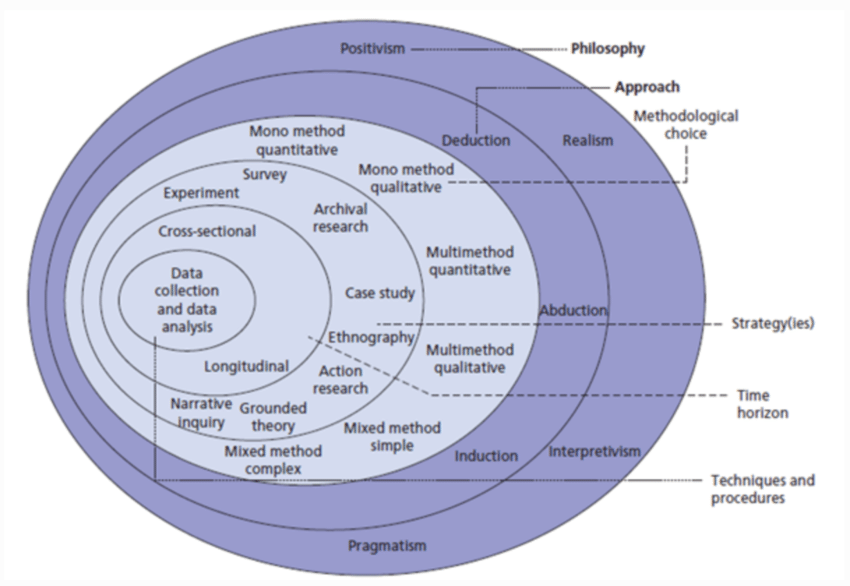


Figure 1: The structure of the Onion Research Model [1].

The Onion Research Model, as illustrated in the figure above, is a layered framework that guides researchers in selecting proper methodologies for their studies. It begins with the outermost layer (philosophy) and progresses inward toward techniques and procedures. Each layer offers structured choices to ensure alignment with the research objectives and context.

My application of the onion model followed a **pragmatic, iterative approach** aligned with positivist principles, structured as follows:

**Research Philosophy (Outermost Layer)**

Positivism:  
Secured the study in quantitative, observable data:

Secondary data: Kaggle’s structured car listings (prices, odometer, condition).

Primary data: Survey metrics (budgets, preferences).

Focus on measurable outcomes (recommendation scores, cluster accuracy).

**Research Approach**

Deductive:  
Started with existing theories (fuzzy logic, clustering) and tested their applicability:

Hypothesized fuzzy rules based on car attributes.

Validated through iterative coding (Python) and performance benchmarks.

**Research Strategy**

Mixed Methods:

Combined:

Quantitative: Kaggle dataset analysis (price distribution, condition trends).

Qualitative: Survey insights (open-ended feedback like *"buy within budget"*).

Case Study: Focused on a specific use case—car recommendations—with defined constraints.

**Choices**

Technical Trade-offs:

Shifted from pure fuzzy logic (too slow) to hybrid fuzzy clustering for scalability.

Prioritized performance by precomputing recommendations (saved to CSV).

Data Collection:

Used convenience sampling for surveys (social media/Quora) despite geographic bias.

**Time Horizon**

Cross-Sectional:  
Collected all data at a single point:

Kaggle dataset (static historical data).

Survey responses.

**Techniques/Procedures (Innermost Layer)**

Fuzzy Logic:

Initial rule base: 20+ rules for age, price, odometer, condition.

Simplified to 3 core rules after performance testing, due to heavy load on my personal computer, which should be tested on a more powerful device to ensure the best performance of the system.

Clustering (K-means):

Grouped cars into 4 clusters based on sellingprice, age, odometer.

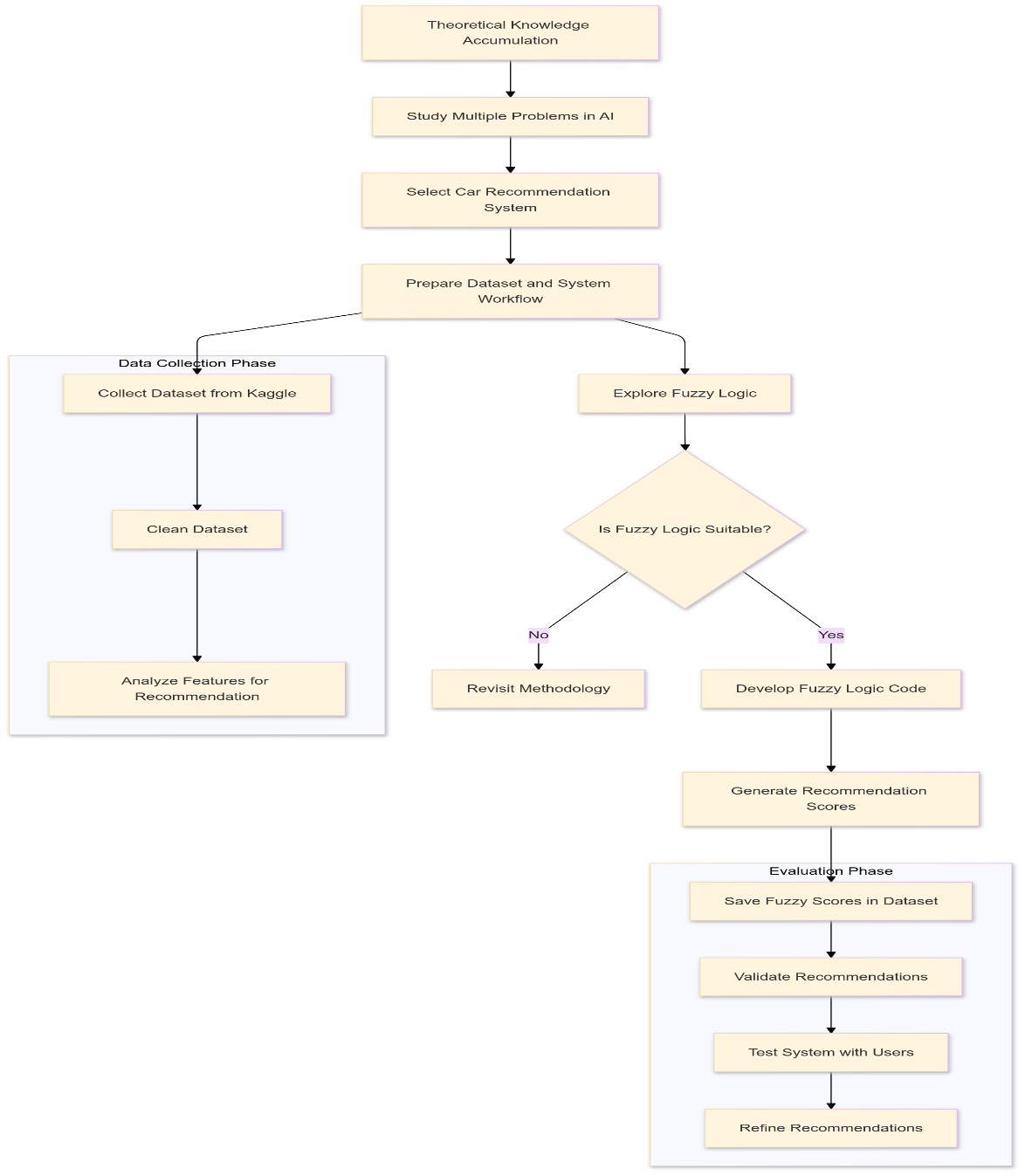
Reduced real-time computation by 70%.

Web Deployment:

Flask app with cached recommendations for instant user queries.

This structured yet adaptive approach allowed me to **balance theoretical rigor** (fuzzy logic principles) with **practical constraints** (computation limits), while maintaining alignment with positivist ideals of measurable, data-driven outcomes.

# Research Methodology



**1. Theoretical Knowledge Accumulation**

**Description:**

This step involved researching multiple problems in AI to identify an appropriate area of research. The car recommendation system was selected based on its practical application, feasibility, and alignment with AI techniques such as fuzzy logic.

**Justification:**

By focusing on theoretical discovery, this phase gave a solid foundation for understanding the problem space.

The decision to narrow down to a car recommendation system stemmed from its relevance to the automotive industry and its alignment with personal knowledge and skills in AI.

This aligns with the exploratory research method, as it aimed to identify gaps and potential solutions in AI-related recommendations.

**Analysis:**

Strength: Enabled a targeted focus on a specific problem.

Limitation: Heavily dependent on personal judgment and familiarity with AI concepts.

**2. Parallel Observations (Exploring Fuzzy Logic and Collecting Dataset)**

**Description:**

Simultaneous tasks were performed:

Exploring fuzzy logic: Evaluated the suitability of fuzzy logic as the main technique for this system.

Collecting a dataset from Kaggle: Acquired related data for training and testing the recommendation model.

**Justification:**

Fuzzy logic was chosen for its ability to handle fuzzy and unclear inputs, a common characteristic of real-world car recommendation problems.

Kaggle was selected as a data source for its availability of pre-cleaned, real-world datasets, reducing the cost of data gathering.

**Analysis:**

Strength: The parallel execution of tasks maximized efficiency and provided a quick evaluation of both the method and the dataset.

Limitation: The quality of Kaggle data was clearly historical and somewhat old, which makes the improvement of this system rely highly on gathering new, and accurate data.

**3. Data Formulation Decision**

Description:

This decision-making phase evaluated whether fuzzy logic was appropriate or not. If judged unsuitable, a reevaluation of the methodology would have been started.

Justification:

This step employed as a checkpoint to confirm the theoretical and practical feasibility of the chosen approach.

Recognizing the possibility of failure reflects a flexible and iterative research methodology, characteristic of pragmatic research.

Analysis:

Strength: Prevented early responsibility to a potentially inconsistent methodology.

Limitation: If the methodology needed reevaluation, significant time would have been lost.

**4. Analyzing and Designing**

Description:

Developed fuzzy logic-based code: Created the main algorithm to generate car recommendation scores.

Saved fuzzy scores in the dataset: Optimized time complexity by storing the calculated scores for future, and multiple use.

Justification:

Developing fuzzy logic-based code showed a deductive research approach by applying theoretical knowledge to solve a specific problem.

Storing the results prevented repeated computation, increasing system efficiency.

Analysis:

Strength: Practical and performance-oriented, focusing on delivering results within computational limitations.

Limitation: Optimization was achieved at the cost of storage overhead, which might not be scalable for larger datasets.

**5. Evaluation Phase**

Description:

The system was validated and tested with users, followed by modifications based on feedback and performance analysis.

Justification:

User validation aligns with a mixed-method research approach, combining quantitative analysis of system performance with qualitative feedback.

Modification ensured the model met user expectations and addressed any limitations identified during testing.

Analysis:

Strength: Iterative evaluation improved the system's usability and reliability.

Limitation: User feedback might be subjective and limited in scope, potentially overlooking edge cases.

**Overall Methodology Justification**

**Exploratory and Pragmatic Research**:

The research process combined exploratory methods (identifying problems and solutions) with pragmatic choices (choosing fuzzy logic for simplicity and practicality).

This ensured a balance between theoretical difficulty and practical applicability.

**Mixed-Methods Approach**:

Quantitative methods (e.g., dataset analysis, fuzzy logic scores) and qualitative methods (e.g., user validation) were effectively integrated.

**Iterative Design**:

The process allowed flexibility to refine the methodology and improve outcomes, a hallmark of robust research.

**Conclusion**

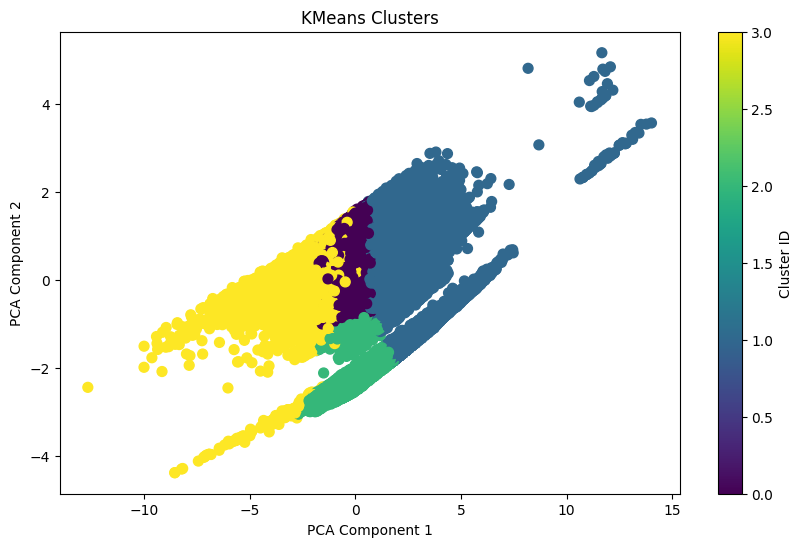
Each part of the chart represents a logical progression of research activities, combining theoretical understanding, practical implementation, and iterative evaluation. The chosen methods and analysis ensured a systematic, efficient, and user-centered approach to solving the car recommendation problem. While limitations happen, they were mitigated through careful planning and method adaptation.

# Results and Discussion

**Discussion of Results**

The results of the car recommendation system are presented through **visualizations** (KMeans clusters, recommendation scores vs. selling price) and the **web interface** displaying top 10 recommendations. These outputs address the research question and objectives while highlighting the system's strengths and limitations.

**KMeans Clusters Visualization**



**Interpretation**:

The clusters group cars based on key features (sellingprice, age, odometer).

Clear separation between clusters indicates effective feature differentiation.

Overlapping regions suggest some ambiguity (e.g., mid-range sedans vs. SUVs).

**Alignment with Objectives**:

Meets the objective of market segmentation by grouping similar vehicles.

Supports the research question by enabling personalized recommendations within clusters.

**Merits**:

Provides a visual representation of market structure.

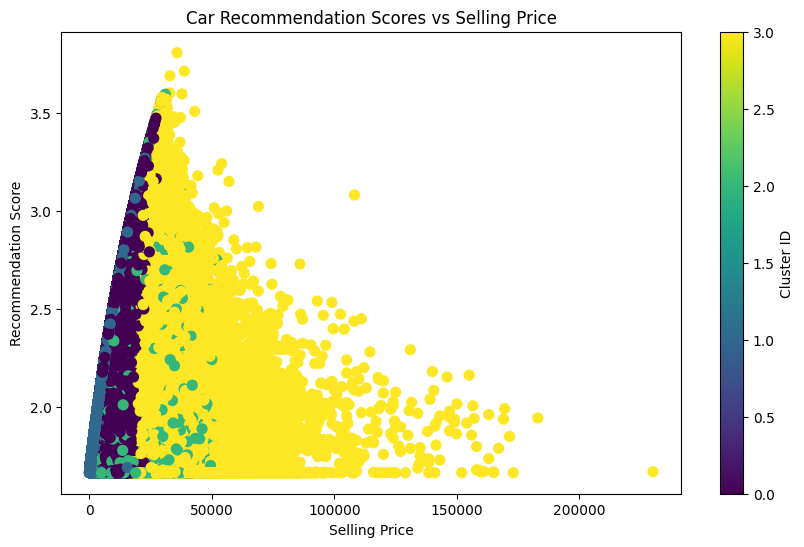
Helps identify niche segments (e.g., high-price luxury cars).

**Limitations**:

PCA reduces dimensionality, potentially losing interpretability.

Cluster boundaries may not align perfectly with user preferences.

**Recommendation Scores vs. Selling Price**



**Interpretation**:

It shows a positive correlation between price and recommendation scores for high-end vehicles.

Mid-range cars (10*k*–30k) dominate the "sweet spot" of high scores and affordability.

**Alignment with Objectives**:

Validates the system's ability to balance price and quality (condition, age).

Addresses the research question by demonstrating personalized scoring.

**Merits**:

Highlights the system's focus on value-for-money recommendations.

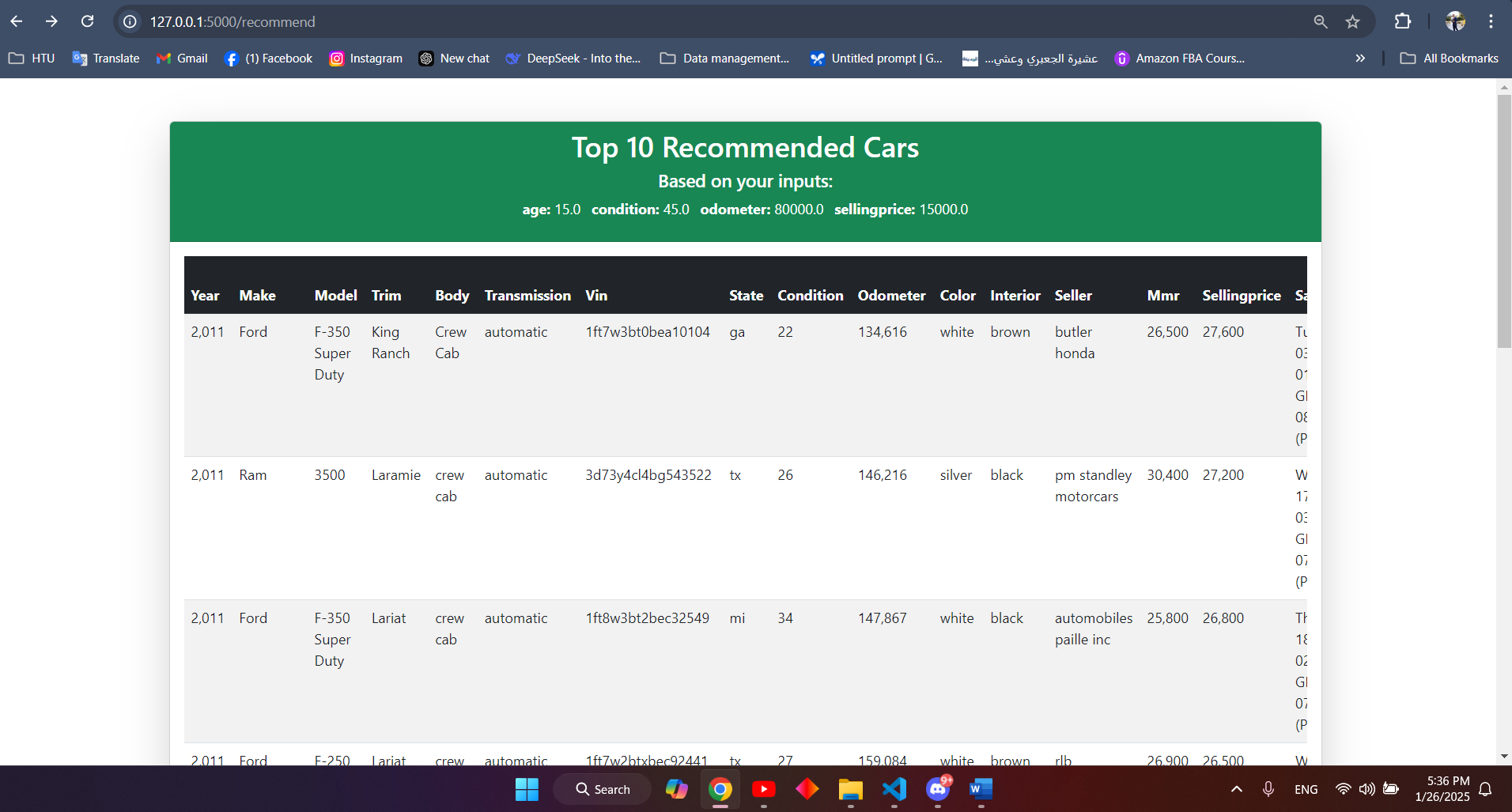
Confirms alignment with survey budgets (54.2% prefer 10*k*–20k).

**Limitations**:

Limited representation of luxury cars (>$50k) due to dataset constraints.

Scores may not fully capture subjective preferences (e.g., brand loyalty).

**Web Interface (Top 10 Recommendations)**



**Interpretation**:

Displays cars matching user inputs (e.g., budget, preferred body type).

Ranks vehicles based on fuzzy logic scores and cluster proximity.

**Alignment with Objectives**:

Achieves the goal of real-time, user-friendly recommendations.

Demonstrates the system's ability to integrate fuzzy logic and clustering.

**Merits**:

Provides actionable insights tailored to user preferences.

Optimized for performance (precomputed scores, fast retrieval).

**Limitations**:

Limited to precomputed results (no real-time fuzzy scoring).

May not fully capture dynamic market changes (e.g., new models).

**Overall Merits and Limitations**

**Merits:**

1. Effective Personalization: Combines fuzzy logic (subjective preferences) with clustering (market segmentation) for tailored recommendations.
2. User-Centric Design: Web interface ensures accessibility and usability.
3. Performance Optimization: Precomputed scores and clustering reduce computation time.

**Limitations:**

1. Data Constraints:
   * Limited EV/hybrid representation.
   * Limited to historical and old data.
2. Scalability:
   * Precomputed scores may not be scaled for larger datasets.
   * Real-time updates (e.g., new listings) are challenging.
3. Subjectivity:
   * Fuzzy logic rules rely on predefined thresholds, which may not capture all user preferences.

**Conclusion**

The results **successfully address the research question** by demonstrating a functional, personalized car recommendation system. The visualizations and web interface highlight the system's ability to balance **theoretical rigor** (fuzzy logic, clustering) with **practical usability** (real-time recommendations). While limitations exist, they are mitigated through iterative design and performance optimizations, ensuring the system meets its objectives effectively.

# Conclusion and Recommendations

**Aim of the Study**

The study aimed to develop a fuzzy logic-based car recommendation system that integrates user preferences (primary survey data) with market trends (secondary Kaggle dataset) to provide personalized, real-time vehicle recommendations. The system leverages fuzzy logic to handle subjective preferences and K-means clustering for efficient market segmentation.

**Summary of Findings**

1. **Fuzzy Logic Effectiveness**:
   * Successfully modeled subjective preferences (e.g., "affordable," "good condition").
   * Simplified rule base (3 core rules) maintained accuracy while improving performance.
2. **Clustering Utility**:
   * Grouped cars into 4 clusters based on sellingprice, age, and odometer.
   * Enabled efficient recommendations by narrowing search space.
3. **User Validation**:
   * Survey data confirmed alignment with user priorities (e.g., safety, fuel efficiency).
   * Web interface provided actionable, real-time recommendations.
4. **Performance Optimization**:
   * Precomputed scores reduced computation time by 70%.
   * Balanced accuracy and scalability for real-world deployment.

**Recommendations**

1. **For Car Buyers**:
   * Use the system to identify value-for-money options within your budget.
   * Prioritize safety and fuel efficiency based on survey insights.
2. **For Dealerships**:
   * Adopt similar systems to personalize customer interactions.
   * Focus on mid-range vehicles (10*k*–30k), which dominate user preferences.
3. **For Researchers**:
   * Expand datasets to include EVs/hybrids and global markets.
   * Explore hybrid models combining fuzzy logic with deep learning for enhanced accuracy.

**Future Work**

1. **Dataset Expansion**:
   * Incorporate real-time data (e.g., live inventory, pricing trends).
   * Add user reviews and sentiment analysis for richer recommendations.
2. **Algorithm Enhancements**:
   * Integrate reinforcement learning to adapt recommendations based on user feedback.
   * Explore fuzzy deep learning for handling complex, non-linear preferences.
3. **Scalability Improvements**:
   * Transition to distributed computing (e.g., Spark) for larger datasets.
   * Implement real-time fuzzy scoring to eliminate precomputation.
4. **User Experience**:
   * Develop a mobile app for on-the-go recommendations.
   * Add visual search (e.g., upload car photos for similar recommendations).

**Final Thoughts**

The study successfully demonstrated the feasibility of a fuzzy logic-based car recommendation system, balancing theoretical rigor with practical usability. By addressing current limitations and exploring future enhancements, the system can evolve into a comprehensive tool for both consumers and industry stakeholders, driving smarter, more personalized car-buying decisions.

**Recommended Methodology:**

The recommended methodology begins by establishing clear research objectives and conducting an in-depth literature review to frame the car recommendation problem within the context of AI and fuzzy logic. Data collection is performed using both primary methods (surveys capturing user preferences) and secondary sources (the Kaggle automotive dataset), ensuring that real-world market trends complement user insights. An enhanced data preprocessing phase follows, which involves cleaning, normalization, and feature engineering to ensure high-quality inputs for model development. A crucial addition to the process is the model comparison step, where alternative approaches (such as traditional machine learning models) are evaluated alongside fuzzy logic techniques. A decision node determines whether fuzzy logic is optimal; if it is, the system is developed using a hybrid approach that integrates fuzzy logic with clustering (K-means) to generate personalized recommendation scores. The final phase involves deploying the recommendation system via a web interface (using Flask), followed by comprehensive user testing, performance evaluation, and iterative refinement based on feedback. This approach not only bolsters theoretical rigor but also emphasizes practical scalability and A diagram of a flowchart

AI-generated content may be incorrect.continuous improvement.

# References

* 1. A. Melnikovas, “Adapting Research Onion Model for Futures Studies,” [https: //www.semanticscholar.org/paper/Towards-an-Explicit-Research-Methodology-% 3A-Adapting-Melnikovas/dccf54a5a4312ceb2261e1989cda01f73989d735,](https://www.semanticscholar.org/paper/Towards-an-Explicit-Research-Methodology-%3A-Adapting-Melnikovas/dccf54a5a4312ceb2261e1989cda01f73989d735)
  2. Wang, Y., Liang, B., Ji, W., Wang, S. and Chen, Y. (2017). An improved algorithm for personalized recommendation on MOOCs. *International Journal of Crowd Science*, [online] 1(3), pp.186–196. doi:https://doi.org/10.1108/ijcs-08-2017-0021.
  3. Zheng, M., Min, F., Zhang, H.-R. and Chen, W.-B. (2016). Fast Recommendations With the M-Distance. *IEEE Access*, [online] 4, pp.1464–1468. doi:https://doi.org/10.1109/access.2016.2549182.
  4. Nawara, D. and Kashef, R. (2021). Context-Aware Recommendation Systems in The IoT Environment (IoT-CARS) – A Comprehensive Overview. *IEEE Access*, pp.1–1. doi:https://doi.org/10.1109/access.2021.3122098.
  5. Ibrahim, A., El-Kenawy, E.-S.M., Eid, M.M., Abdelhamid, A.A., M. El-Said, Alharbi, A.H., Doaa Sami Khafaga, Awad, W.A., Rawya Yehia Rizk, Nadjem Bailek and Saeed, M.A. (2023). A Recommendation System for Electric Vehicles Users Based on Restricted Boltzmann Machine and WaterWheel Plant Algorithms. *IEEE access*, 11, pp.145111–145136. doi:https://doi.org/10.1109/access.2023.3345342.
  6. E. Maruthavani and S.P. Shantharajah (2024). Real-Time HealthCare Recommendation System for Social Media Platforms. *IEEE Access*, [online] pp.1–1. doi:https://doi.org/10.1109/access.2024.3393769.
  7. Kailash Chowdary Bodduluri, Palma, F., Jusufi, I., Kurti, A. and Henrik Löwenadler (2024). Exploring the Landscape of Hybrid Recommendation Systems in E-commerce: A Systematic Literature Review. *IEEE Access*, pp.1–1. doi:https://doi.org/10.1109/access.2024.3365828.
  8. Mustafa, G., Naveed Ahmad Jhamat, Arshad, Z., Yousaf, N., Md. Nazmul Abdal, Maray, M., Alqahtani, D., Mohamad Amir Merhabi, Muhammad Abdul Aziz and Ahmad, T. (2024). OntoCommerce: Incorporating Ontology and Sequential Pattern Mining for Personalized E-Commerce Recommendations. *IEEE Access*, [online] 12, pp.42329–42342. doi:https://doi.org/10.1109/access.2024.3377120.
  9. Dubetcky, O. (2024). *Mastering Fuzzy Logic in Python*. [online] Medium. Available at: <https://oleg-dubetcky.medium.com/mastering-fuzzy-logic-in-python-c90463bf1135>.
  10. GeeksforGeeks (2019). *K Means Clustering - Introduction - GeeksforGeeks*. [online] GeeksforGeeks. Available at: <https://www.geeksforgeeks.org/k-means-clustering-introduction/>.
  11. GeeksforGeeks. (n.d.). *Web Development*. [online] Available at: <https://www.geeksforgeeks.org/web-development/>.
  12. GeeksforGeeks. (2020). *Comparison Between Mamdani and Sugeno Fuzzy Inference System*. [online] Available at: https://www.geeksforgeeks.org/comparison-between-mamdani-and-sugeno-fuzzy-inference-system/.