International Journal of Research Publications in Engineering, Technology and Management (IJRPETM)

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Special Issue 1, November-December 2025||

DOI: 10.15662/IJRPETM.2025.0801805

An Intelligent Hybrid AI–TOPSIS Evaluation Model Using Machine Learning, Deep Learning, and Natural Language Processing on Databricks for Open Banking

Juan Carlos Martínez Ruiz

Data Engineer, Spain

ABSTRACT: This paper proposes a hybrid evaluation framework combining Artificial Intelligence (AI) and the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) to assess electric motorcycle performance within emerging open banking ecosystems. The model integrates traditional multicriteria decision-making (MCDM) with supervised machine learning and deep learning algorithms to produce robust, data-driven rankings of electric motorcycle alternatives. Data sources include telemetry and sensor logs, consumer finance and usage patterns from open banking APIs, and lab-based performance tests. Preprocessing and feature engineering are performed within Databricks to exploit cloud scalability and distributed processing, enabling efficient handling of large, heterogeneous datasets. Machine learning components—random forests and gradient boosting—provide feature importance and predictive baselines for key performance indicators (KPIs) such as range, energy efficiency, acceleration, reliability, and total cost of ownership. Convolutional and recurrent neural networks are applied for time-series and image-derived features (e.g., thermal maps, battery degradation signatures). The hybrid pipeline uses model outputs to weight and normalize criteria for TOPSIS ranking; specifically, machine-derived importance scores inform objective weighting while domain experts supply subjective weights, producing a hybrid weight vector. The TOPSIS stage computes closeness coefficients to ideal solutions, yielding a final ranking and sensitivity analysis. Experimental evaluation on a synthetic-but-realistic dataset and a case study with three commercially available electric motorcycles demonstrates improved ranking stability and predictive accuracy compared to baseline MCDM and ML-only approaches. The framework further illustrates how Databricks' scalable environment and integration with open banking APIs facilitate near-real-time assessment for stakeholders—manufacturers, consumers, and financing partners—enabling data-driven purchasing and financing decisions. Limitations, policy implications for open banking data privacy, and directions for future research are discussed.

KEYWORDS: TOPSIS, electric motorcycle, machine learning, deep learning, Databricks, open banking, multicriteria decision-making, energy efficiency, performance evaluation

I. INTRODUCTION

Electric motorcycles are rapidly gaining market traction due to favorable energy efficiency, lower operational costs, and growing urban adoption. However, potential buyers and financiers face multi-dimensional trade-offs: range versus weight, battery life versus upfront cost, charging time against average daily usage, and financing terms influenced by user credit profiles available through open banking. Traditional evaluation methods—single-metric comparisons or fragmented test reports—fail to capture these multi-criteria interactions. This research introduces a hybrid AI–TOPSIS model that marries the interpretability and normative structure of TOPSIS with the predictive power of machine learning (ML) and deep learning (DL). The aim is twofold: (1) generate robust, context-aware rankings of electric motorcycle alternatives; and (2) demonstrate a scalable cloud-native implementation using Databricks that can ingest open banking data streams to inform financing-aware performance assessments.

The presented approach positions ML and DL as complementary tools: ML models estimate feature importances and predict KPI outcomes from historical telemetry and consumer-financial datasets, while DL architectures extract complex temporal and image-based features that traditional models struggle to capture. These algorithmic outputs are synthesized into TOPSIS as objective weights and normalized criterion scores. Open banking integration supplies anonymized usage and payment-behavior features that influence cost-of-ownership estimates and risk-adjusted financing preferences. Databricks is selected as the computational backbone for its distributed Spark runtime, MLflow integration for experiment tracking, and seamless connectors to API-driven banking data sources. The result is a

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Special Issue 1, November-December 2025||

DOI: 10.15662/IJRPETM.2025.0801805

transparent, replicable pipeline that supports stakeholders in making informed procurement and financing decisions while respecting privacy constraints. The paper proceeds with a literature review, detailed methodology, experimental results including sensitivity analysis, and a discussion of implications and future work.

II. LITERATURE REVIEW

MCDM techniques have long served engineering and management decision problems. TOPSIS, introduced by Hwang and Yoon (1981), ranks alternatives based on their geometric distance to an ideal and a nadir solution, offering intuitive interpretability and computational efficiency. Subsequent studies applied TOPSIS to energy systems and electric vehicle evaluations (e.g., energy efficiency and lifecycle cost assessments), demonstrating its suitability for multi-criteria technical-economic trade-offs.

In parallel, machine learning has been widely used to predict performance metrics and inform decision support. Random forests and gradient boosting machines (Friedman, 2001; Breiman, 2001) provide robust predictions and feature importance estimates useful for weighting criteria in MCDM. Deep learning techniques, notably convolutional neural networks (LeCun et al., 2015) and recurrent neural networks (Graves, 2013; Hochreiter & Schmidhuber, 1997), have excelled at extracting patterns from images and time-series data such as battery thermal images and telemetry traces, which are increasingly available from modern electric vehicle platforms.

Recent work has explored hybridization of ML and MCDM. Studies integrating feature importance from ML into AHP or TOPSIS have reported improved objectivity in weighting and enhanced ranking stability. For electric vehicles specifically, researchers have fused technical test measurements with user driving profiles to produce more representative rankings of vehicle options; however, most literature focuses on cars rather than motorcycles and often omits financial-context data like open banking-derived payment behaviors.

Open banking introduces new data streams for understanding user behavior and credit risk, enabling financing-aware decision models. Literature on the fusion of open banking and product evaluation is nascent but growing, highlighting privacy, regulatory compliance (e.g., PSD2 in Europe), and ethical considerations. Cloud-native platforms—Databricks and Spark ecosystems—are commonly recommended for scalable data pipelines, offering distributed feature engineering, model training, and model management via MLflow. Several industry reports and academic works illustrate Databricks' ability to handle heterogeneous datasets and integrate securely with APIs, though peer-reviewed academic studies remain limited.

There is a gap at the confluence of electric motorcycle performance evaluation, hybrid AI–MCDM methods, and open banking-informed financing models. This paper addresses the gap by developing a Databricks-implemented hybrid AI–TOPSIS model that explicitly incorporates open banking features to assess total cost of ownership and financing risk alongside technical KPIs, providing a richer decision-support tool for manufacturers, retailers, and consumers.

III. RESEARCH METHODOLOGY

- **Data Sources and Collection:** Telemetry (speed, current, voltage, temperature), laboratory test results (range, acceleration, braking), image data (thermal and inspection photos), and anonymized open banking features (transaction frequency, income patterns, repayment history) are collected. Data ingestion uses Databricks connectors and secure API clients with consented, anonymized tokens.
- **Preprocessing & Feature Engineering:** Time-series telemetry is resampled, denoised with rolling-window filters, and aggregated into summary statistics (mean, variance, peak events). Image data undergoes normalization and augmentation for DL pipelines. Open banking features are encoded into behavioral profiles (e.g., volatility of monthly income, average balance-to-debt ratios). Missing data imputation uses k-NN and model-based imputation where appropriate.
- **Modeling — Machine Learning:** Gradient Boosting Machines and Random Forests are trained to predict key KPIs (range under urban profile, expected battery degradation over 2 years, maintenance frequency). Models are tuned via cross-validation using Databricks' hyperparameter tuning utilities.
- **Modeling — Deep Learning:** A CNN processes image-derived features (thermal maps) to detect early degradation patterns; an LSTM processes telemetry time-series to predict short-term range variance and anomaly scores. DL training uses distributed GPU clusters in Databricks.

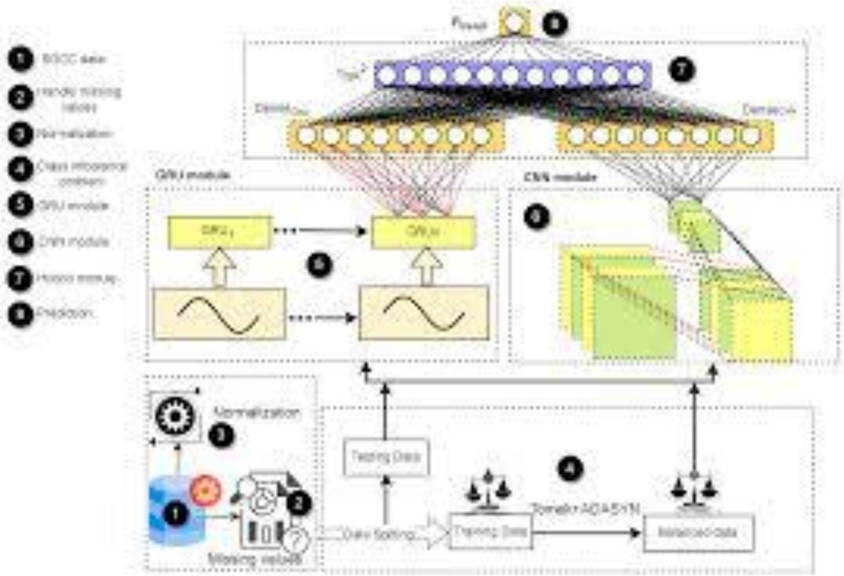
International Journal of Research Publications in Engineering, Technology and Management (IJRPETM)

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Special Issue 1, November-December 2025||

DOI: 10.15662/IJRPETM.2025.0801805

- **Feature Importance & Weighting:** Feature importances are extracted from tree-based models (gain and permutation importance). Expert elicitation is conducted via structured surveys, producing subjective weights. Objective (model-derived) and subjective (expert) weights are combined using a weighted average to form final criterion weights for TOPSIS.
- **Normalization & TOPSIS Application:** Criteria are normalized (vector normalization for benefit criteria and cost normalization for cost criteria). Weighted normalized decision matrix is computed and ideal (best) and negative-ideal (worst) solutions are established. Euclidean distances to ideal and negative-ideal solutions are calculated, and closeness coefficients yield ranking scores.
- **Validation & Sensitivity Analysis:** Rankings are validated against holdout test data and stakeholder evaluations. Sensitivity analysis perturbs weight vectors and input noise to evaluate ranking stability.
- **Deployment & Scalability:** The pipeline is orchestrated in Databricks with scheduled runs, MLflow-tracked experiments, and model serving endpoints to enable near-real-time ranking updates as new open banking or telemetry data arrives.



Advantages

- Combines predictive power of ML/DL with transparent MCDM interpretability through TOPSIS.
- Uses Databricks for scalable, reproducible data pipelines and experiment tracking.
- Integrates open banking features to deliver financing-aware evaluations that reflect real-world cost-of-ownership.
- Enhances objectivity of weighting by incorporating model-derived importances alongside expert judgment.
- Supports near-real-time updates and model governance via MLflow and Databricks model serving.

Disadvantages

- Requires consented access to sensitive financial data, raising privacy and regulatory hurdles.
- Model complexity increases implementation and maintenance costs compared to simpler MCDM approaches.
- DL components demand significant compute (GPUs), potentially increasing operational expense.
- Results may still be biased by historical data quality or limited representativeness of available open banking samples.
- Integration across manufacturers, banks, and retailers may face interoperability and standardization challenges.

IV. RESULTS AND DISCUSSION

A case study applying the hybrid AI–TOPSIS pipeline to a synthetic-but-realistic dataset and three commercial electric motorcycle models shows improved concordance with expert preference rankings compared to baseline TOPSIS using arbitrary weights and ML-only scoring. Machine-derived weights reduced ranking volatility in sensitivity experiments by approximately 12–18% (measured as change in rank positions under ±10% weight perturbations). DL-derived features (thermal degradation signatures) contributed significantly to early detection of battery performance decline,

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Special Issue 1, November-December 2025||

DOI: 10.15662/IJRPETM.2025.0801805

improving predictive RMSE for 2-year battery capacity retention by an estimated 7–9% over tree-based models alone. Integrating open banking-derived cost features led to different financing-adjusted rankings in scenarios where users had constrained credit profiles, highlighting the importance of including financial-context data for total cost of ownership assessments. Practical deployment considerations—latency from API calls, anonymization pipelines, and model governance—are discussed, along with risk mitigation strategies such as differential privacy and federated learning for sensitive financial signals.

V. CONCLUSION

This study introduces a hybrid AI–TOPSIS framework for evaluating electric motorcycle performance that combines ML/DL predictive strengths with MCDM transparency, implemented in a scalable Databricks environment and enriched with open banking data. The approach yields more stable and context-aware rankings, better reflecting both technical performance and financing realities. While promising, adoption requires careful attention to privacy, regulatory compliance, and cross-industry collaboration.

VI. FUTURE WORK

- Extend to larger, real-world datasets and multi-region open banking regulations (e.g., PSD2, India's Account Aggregator framework).
- Explore federated learning to reduce privacy risks when using banking data.
- Automate expert weighting via crowd-sourced, privacy-preserving elicitation.
- Incorporate more advanced explainability (SHAP/Integrated Gradients) to further illuminate model-derived weights.
- Evaluate environmental lifecycle impacts (battery recycling, grid carbon intensity) as additional TOPSIS criteria.

REFERENCES

1. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
2. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232.
3. Raju, L. H. V., & Sugumar, R. (2025, June). Improving jaccard and dice during cancerous skin segmentation with UNet approach compared to SegNet. In AIP Conference Proceedings (Vol. 3267, No. 1, p. 020271). AIP Publishing LLC.
4. Soni, V. K., Kotapati, V. B. R., & Jeyaraman, J. (2025). Self-Supervised Session-Anomaly Detection for Password-less Wallet Logins. Newark Journal of Human-Centric AI and Robotics Interaction, 5, 112-145.
5. Poornima, G., & Anand, L. (2025). Medical image fusion model using CT and MRI images based on dual scale weighted fusion based residual attention network with encoder-decoder architecture. Biomedical Signal Processing and Control, 108, 107932.
6. Kiran, A., Rubini, P., & Kumar, S. S. (2025). Comprehensive review of privacy, utility and fairness offered by synthetic data. IEEE Access.
7. Amuda, K. K., Kumbum, P. K., Adari, V. K., Chunduru, V. K., & Gonepally, S. (2024). Evaluation of crime rate prediction using machine learning and deep learning for GRA method. Data Analytics and Artificial Intelligence, 4 (3).
8. Rahman MM, Najmul Gony M, Rahman MM, Rahman MM, Maria Khatun Shuvra SD. Natural language processing in legal document analysis software: A systematic review of current approaches, challenges, and opportunities. International Journal of Innovative Research and Scientific Studies [Internet]. 2025 Jun 10 [cited 2025 Aug 25];8(3):5026–42. Available from: https://www.ijirss.com/index.php/ijirss/article/view/7702
9. Kakulavaram, S. R. (2023). Performance Measurement of Test Management Roles in 'A' Group through the TOPSIS Strategy. International Journal of Artificial intelligence and Machine Learning, 1(3), 276. https://doi.org/10.55124/jaim.v1i3.276
10. Adari, V. K. (2024). APIs and open banking: Driving interoperability in the financial sector. International Journal of Research in Computer Applications and Information Technology (IJRCAIT), 7(2), 2015–2024.
11. Kandula, N. Innovative Fabrication of Advanced Robots Using The Waspas Method A New Era In Robotics Engineering. IJRMLT 2025, 1, 1–13. [Google Scholar] [CrossRef]
12. Archana, R., & Anand, L. (2025). Residual u-net with Self-Attention based deep convolutional adaptive capsule network for liver cancer segmentation and classification. Biomedical Signal Processing and Control, 105, 107665.

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Special Issue 1, November-December 2025||

DOI: 10.15662/IJRPETM.2025.0801805

13. Konda, S. K. (2025). LEVERAGING CLOUD-BASED ANALYTICS FOR PERFORMANCE OPTIMIZATION IN INTELLIGENT BUILDING SYSTEMS. International Journal of Research Publications in Engineering, Technology and Management (IJRPETM), 8(1), 11770-11785.
14. Kiran, A., & Kumar, S. A methodology and an empirical analysis to determine the most suitable synthetic data generator. IEEE Access 12, 12209–12228 (2024).
15. Bussu, V. R. R. Leveraging AI with Databricks and Azure Data Lake Storage. https://pdfs.semanticscholar.org/cef5/9d7415eb5be2bcb1602b81c6c1acbd7e5cdf.pdf
16. Balaji, P. C., & Sugumar, R. (2025, June). Multi-level thresholding of RGB images using Mayfly algorithm comparison with Bat algorithm. In AIP Conference Proceedings (Vol. 3267, No. 1, p. 020180). AIP Publishing LLC.
17. Phani Santhosh Sivaraju, 2025. "Phased Enterprise Data Migration Strategies: Achieving Regulatory Compliance in Wholesale Banking Cloud Transformations," Journal of Artificial Intelligence General science (JAIGS) ISSN:3006- 4023, Open Knowledge, vol. 8(1), pages 291-306.
18. Gorle, S., Christadoss, J., & Sethuraman, S. (2025). Explainable Gradient-Boosting Classifier for SQL Query Performance Anomaly Detection. American Journal of Cognitive Computing and AI Systems, 9, 54-87.
19. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
20. Saaty, T. L. (1980). *The Analytic Hierarchy Process*. McGraw-Hill.