International Journal of Academic Engineering Research (IJAER)
ISSN: 2643-9085
Vol. 9 Issue 5 May - 2025, Pages: 26-34

Integration of Demographic, Behavioral, and Contextual Data in AutoGluon for Holistic Buyer Modeling

Mayowa Alonge

Abstract: *With the era of data-driven decision-making, consumer behavior modeling requires the fusion of various data modalities for an exhaustive buyer profile. This research explores the incorporation of demographic, behavioral, and contextual data in AutoGluon, an open-source AutoML library, with the vision of developing an integrated buyer modeling framework. We present a pipeline that preprocesses heterogenous data sources, imputes missing values, and encodes complex feature interactions to train high-performance predictive models with minimal human tuning. With multi-modal ensemble learning and feature importance of AutoGluon, the framework facilitates robust buyer segmentation and propensity scoring. Empirical experiments demonstrate significant performance gains over unimodal models, particularly in dynamic e-commerce and personalized marketing settings. The findings demonstrate the value of contextual enrichment and behavior-demographic synergy in predictive buyer modeling, with specific emphasis on AutoGluon's effectiveness for large-scale deployment of AutoML for business intelligence in the real world.*

Keywords*: AutoGluon, demographic data, behavioral data, contextual data, buyer modeling, machine learning, AutoML, consumer behavior, personalization, recommendation systems, feature engineering, ensemble learning, targeted advertising.*

Introduction

1.1 Purpose and Scope

Buyer modeling is the art and science of developing forecasting models that understand, group, and forecast customer behavior and decision-making. It is applied to personalization and ethical marketing because it enables companies to personalize product recommendations, message strategy, and engagement mechanisms through individual customer profiles. Good buyer modeling maximizes customer satisfaction and conversion and marketing ROI. Historically, buyer profiles have been constructed from isolated streams of data such as demographic or transactional. However, in more sophisticated digital landscapes, the combination of behavior (i.e., clickstream, browsing history) and context (e.g., time, device, location) is necessary to construct a full view of the buyer. Combining these extremely different forms of data allows for more advanced segmentation, improved predictability, and more behavioral understanding.

This paper explores the integration of demographic, behavioral, and contextual data with AutoGluon, an open-source AutoML library developed by Amazon. The goal is to develop a single, horizontally scalable pipeline that reduces the need for manual feature engineering while generating high-performing models suitable for real-world personalization use cases.

1.2 Why AutoGluon?

AutoGluon is distinct from other AutoML libraries due to its ability to automatically process heterogeneous datasets with little user input. Key advantages are:

• Automated pre-processing of heterogeneous data types (categorical, numerical, text, time-based).
• Stacked ensemble over multiple families of models to deliver stable predictive performance.
• Feature importance insight, so end-users can understand model decisions.
• Ease of use, with minimal lines of code required to train good models.

As compared to other AutoML libraries:

• H2O.ai AutoML is robust but needs further tuning and lacks deep contextual integration out-of-the-box.
• TPOT, being a genetic programming-based, provides interpretability but perhaps may be

computationally expensive and slower for voluminous data.
•DataRobot possesses enterprise-oriented but commercially oriented features that are not suited for educational or low-volume use cases.
AutoGluon's balance of performance, flexibility, and availability makes it highly suitable to mix together heterogeneous buyer-centric data and investigate commercial predictive models at high speeds in perpetually changing commercial environments.

Background

2.1 Types of Buyer Data

Modern buyer modeling relies on the combination of different types of data that identify different elements of a shopper's profile and behavior. The three primary types of data used in this study are:
• Demographic Information: It comprises very stable attributes such as age, sex, income class, education, and geography. Demographics provide preliminary pointers for segmentation which help create buyer personas in addition to predicting preference on a macro level.
• Behavioral Data: These are dynamic indicators that are derived from user behavior, such as clickstream patterns, transaction history, session length, and browsing. Behavioral signals monitor real-time or near-real-time intention and are essential for anticipating instantaneous activities such as purchases or churn.
•Contextual Data: Context provides situational insight to behavior and demographics. Device type, time of day, geolocation, and campaign source features allow marketers to not only know what the buyer is doing, but when and where he or she is doing it. Contextual data introduces added relevance to personalization by translating prediction to environmental or time variables.
All these types of data in and of themselves hold little in the way of clue.
But combined, they enable more insight and more targeted targeting across several customer touchpoints.

2.2 Complexity of Data Merging

Combining these seemingly unrelated data types for machine learning is faced with several non-obvious challenges:
• Lack of Density in Data and Missing Values: The demographics might be sparse, the behavior could be session- or intermittently-logged data, and context signals might not even be logged every single time. Such sparsenesses would result in challenging model training and bias development in its prediction.
•Multiple Data Forms: The data set includes categorical (e.g., gender, device), numerical (e.g., income, session duration), and temporal (e.g., timestamped interactions) features. Efficient encoding and normalization of these multiple forms needs to be achieved so that there is no loss of information and meaning is maintained.
• Feature Interdependence and Interaction Effects: Purchase decisions often arise from complex interactions between variables—a wealthier customer, say, may respond differently to gifts based on the gadget category or time of day. Model such non-linear associations with advanced methods that can learn and employ automatically such interactions without feature engineering.
AutoML toolkits like AutoGluon have embedded solutions to handle most of these integration problems so that practitioners do less strategy and experimentation and more preprocessing and pipeline tuning.

3. Data Preprocessing and Feature Engineering

3.1 Data Collection and Cleaning

The foundation of effective buyer modeling lies in the effective merging and cleaning of data between sources. During this project, demographic, behavior, and context data were collected from various systems such as CRM systems, web analytics, and e-commerce databases. Data integration involved:

• Merging datasets using unique customer IDs or session keys for synchronization across data modalities.

• Handling missing values via appropriate imputation methods (e.g., mean/median for numerical features, mode or "unknown" class for categorical).

• Handling differences such as naming conventions, time zones, or duplicate records using rule-based decision-making or reliable source ranking.

• Removing noise and outliers based on statistical bounds or domain-knowledge-based heuristics in order to avoid model training skew.

This cleaning left us with a robust and uniform dataset, ready to be input into machine learning workflows in an automated manner.

3.2 Encoding Strategies

Since input data is diverse, feature encoding techniques were applied to make the features ML-ready:

• Categorical Encoding: Based on cardinality and model type, we employed:

o One-hot encoding for low-cardinality nominal features like gender or device type.

o Label encoding for ordinal features (e.g., education level).

o Frequency encoding for high-cardinality features like geolocation or product categories.

• Temporal Encoding: Timestamps in behavior logs and context records were decomposed into features like:

Hour of day, day of week, month, and season to pick up seasonal and cyclical effects.

Time since last interaction (optional) to pick up recency-based behavior.

Textual Encoding (if available): Any text features such as user search queries or product descriptions were encoded as:

TF-IDF vectors for sparse, interpretable representations.

Pretrained word embeddings (e.g., GloVe or FastText) for denser semantic capture, especially in models that have support for deep learning.

AutoGluon's automated preprocessing was used wherever possible to reduce these encoding steps with minimal hand-engineering of features.

3.3 Feature Transformation and Selection

To achieve best model performance and interpretability, several transformation and selection techniques were applied:

• Standardization/Normalization: Numerical features like income and session length were standardized to zero-mean and unit variance, especially for scale-sensitive models (e.g., neural networks or SVMs).

• Feature Importance Analysis: Following modeling, we employed:

SHAP (SHapley Additive exPlanations) values to describe per-feature contributions.

Permutation importance for model-agnostic interpretation of the contribution of each variable to performance.

Dimensionality Reduction (if required): In cases where feature space sparsened or became strongly multicollinear, techniques like Principal Component Analysis (PCA) or AutoGluon's internal feature pruning was used to decrease dimensionality without loss of predictive power.

These preprocessing and engineering processes put the dataset in a form suitable for AutoGluon's automated pipeline, improving both efficiency as well as model performance.

4 .AutoGluon Pipeline Configuration

4.1 AutoGluon Dataset Integration

AutoGluon's TabularPredictor supports datasets in structured tabular form. To ensure integration among demographic, behavioral, and contextual information, all the features were unified into a single DataFrame where each row corresponds to a unique user or session and each column a respective feature.

AutoGluon supports multi-modal input if they are tabular in forms. These are:

•\ttNumerical columns (e.g., session length, income).

• Categorical columns (e.g., gender, device type).

• Datetime fields (e.g., timestamp of the last purchase).

• Text columns (if present, e.g., user reviews or queries).

Behavioral data, possibly in time-series or log format to begin with, were aggregated to session-level or user-level metrics (e.g., total clicks, average session length) so that it could be fitted in the tabular model.

4.2 Custom Feature Metadata

While AutoGluon is able to automatically infer feature types, custom feature metadata provides more precise treatment, particularly for edge cases. This is achieved through the feature_metadata parameter to TabularPredictor.fit().

Individual feature types are:

• Numerical: Discrete or continuous quantitative features.

• Categorical: Non-ordinal labels or classes (e.g., campaign source, location).

• Text: Unstructured text fields that can be handled by AutoGluon's NLP models.

• Datetime: Timestamps from which AutoGluon is able to automatically extract temporal components.

Accurate type definitions improve preprocessing decisions (e.g., embedding vs. one-hot encoding), model selection, and performance in general.

4.3 Training and Hyperparameter Tuning

AutoGluon simplifies training models without sacrificing powerful customizations:

•\tTime Constraints: The time_limit parameter constrains model training to a budgeted duration of time, especially helpful in production or quick prototyping settings.

•\tEnsembling Strategies: AutoGluon uses automatically:

o\tStacked ensembling, which combines predictions from a diverse set of base models (e.g., LightGBM, CatBoost, neural nets).

o Bagging as variance reduction technique through training several models on bootstrap samples.

• Hyperparameter Tuning: Even though AutoGluon does some minor tuning internally as well, users may also supply search spaces or employ the hyperparameter_tune_kwargs argument as a way of using more sophisticated optimization through Bayesian or random search habits.

• Dealing with Imbalanced Data: For classification problems involving imbalanced class distributions (e.g., rare purchases), there are various possibilities:

o Utilizing AutoGluon's class_weights parameter.

o\tApplying resampling algorithms (e.g., SMOTE, undersampling) prior to input.

n\tUtilizing proper metrics like AUC or F1-score instead of accuracy for assessment.

These options allow AutoGluon to perform best predictive results with minimal human adjustment, yet also permit domain-specific constraints.

5.Model Evaluation and Interpretability

5.1 PerforMance Metrics

To ensure the simplicity and resilience of the buyer modeling framework, performance was quantified using a collection of efficacy metrics suitable to the type of predictive task:

• Classification Tasks (e.g., probability of purchase prediction, risk of churn):

o AUC (Area Under the ROC Curve): Assesses the ability of the model to distinguish classes, particularly with imbalanced data.

o Accuracy: Indicates overall correctness, but might be misleading when severe class imbalance occurs.

o F1-Score: Harmonic mean of precision and recall; particularly helpful when there are different business impacts of false positives and false negatives.

•Regression Tasks (if being used, e.g., calculating customer lifetime value or level of spending):

oRMSE (Root Mean Squared Error): More punishment for large errors, with more focus on model accuracy in aggregate.

oMAE (Mean Absolute Error): Easier to interpret and less influenced by outliers than RMSE.

AutoGluon also does cross-validation and leaderboard score-based ranking of model performance automatically so that base and ensemble models can be easily compared.

5.2 Interpretability Tools

Model transparency is crucial for high-stakes applications like personalization or campaign targeting. AutoGluon and other tools were used to help with interpretability at global as well as segment levels:

•\tFeature Importance Plots: AutoGluon's integrated predictor.feature_importance() method ranks features based on their contribution to the model's predictions. This identifies the most impactful factors on demographic, behavioral, and contextual dimensions.

•\tSHAP Integration (optional): For finer insight, SHAP (SHapley Additive exPlanations) was applied to chosen models to:

o\tMeasure how much each feature contributes to a given prediction.

o\tVisualize interactions and feature dependencies.

o\tCompare effects across user segments or prediction classes.

• Segment-Specific Insights: Post-hoc analysis was conducted for interpretation purposes of variation in model prediction among subgroups, such as:

Age group (e.g., young users vs. older users).

Acquisition channel (e.g., organic vs. paid channels).

Device type (e.g., mobile vs. desktop).

These interpretability techniques not only contribute to stakeholder trust but also offer insights on marketing strategy through revelation of actionable findings on central drivers of purchase.

Case Study / Example Implementation

6.1 Dataset Description

For the purposes of practical illustration of the presented buyer modeling pipeline, we have used the Kaggle E-Commerce Behavior Data public dataset. It contains anonymized user logs from an e-shop, which consist of:
•Demographic attributes (simulated or inferred): region, age group.
•Behavioral signals: page views, cart additions, product clicks, and transactions over time.
•Context data: session ID, device type, and timestamp.
The range of data was on the order of millions of events across thousands of users and provided a realistic context for exercising personalized recommendation and conversion prediction use cases.
Alternatively, a synthetically created data set of similar structure was built in order to experiment and replicate when public data were limited or domain-specific augmentation was required.

6.2 Step-by-Step Walkthrough

The build-out proceeded as follows:

1.Data Preparation:
Merged logs into a user-level table by session IDs and user IDs.
Engineered session-based aggregates (e.g., view count, cart conversion, session length).
Mapped timestamped to time-based features (e.g., hour of the day, day of the week).
Imputed missing data and encoded categorical variables by frequency and one-hot encoding.

2.AutoGluon Pipeline:
Constructed the TabularDataset and made feature_metadata to specify categorical, numerical, datetime, and text fields.
Installed TabularPredictor with eval_metric='f1' for classification and specified time limitations (e.g., time_limit=3600 seconds).

3. Model Training:
Invoked.fit() using AutoGluon's default ensemble and stacking parameters.
Allowed for hyperparameter tuning automation with random search and early stopping.

4.Evaluation:
Used AutoGluon's leaderboard and.evaluate() functions to evaluate AUC, accuracy, and F1-score.
Did k-fold validation and saved models with best generalization performance.

5.Interpretation:
Generated feature importance plots.
Used SHAP values on top-performing model to explain predictions across user segments.

6.3 Insights Derived

Model analysis and interpretability results revealed:
Drivers of Buyer Behavior:
Number of product views, session length, and cart adds were the most predictive behavioral indicators.
Contextual elements like day of the week and device were highly correlated with conversion likelihood.

Cross-Segment Comparisons:

Users 18–25 were more contextually affected by elements like mobile behavior and evening browsing.
Users >45 showed more significant associations between purchase behavior and demographic stability (e.g., income, location).

Business and Operational Implications:

Insights provide time-efficient targeting (e.g., push notifications at times of highest engagement).
Empresses the value of use of recency of behavior with static attributes to lift conversion models.
Prescribes content strategy by segment, e.g., UI tuning by device and local pricing sensitivity.
This case study illustrates how AutoGluon facilitates an end-to-end buyer modeling workflow—from multi-source data fusion to explainable predictions—allowing marketers and analysts to make data-driven decisions efficiently.

Challenges and Limitations

7.1 Data Limitaciones

While the integration strategy worked, several data-related shortcomings persisted:

•Incomplete User Profiles: A large portion of the user population had missing or incomplete demographic information, such as income or location. This incompleteness can impede the ability to segment and personalize exactly, lowering model accuracy for specific groups of users.

•Noisy Data: The context and behavior data will be noisy since there are inaccuracies in the logging (e.g., missing session information, duplicated entries, or time stamps). The noise creates ambiguity during training, which can affect model performance unless it is handled.

• Anonymized Records: Most e-commerce records of users are anonymized for privacy purposes. Although it protects user identity, it hinders creating complete buyer profiles and appropriate demographic segmentation.

These data restrictions call for emphasis at preprocessing and imputation stages, and some cases require domain knowledge to enhance input data quality.

7.2 AutoML Constraints

AutoGluon has an end-to-end AutoML pipeline but has certain built-in constraints:

• Limited Tunability: Even though AutoGluon is very automated, this does limit those who would want to toy around with high-level, model-specific tuning. For example, even though it comes with support for hyperparameter tuning out-of-the-box, one has rather limited capability to adjust specific phases like feature preprocessing or model choosing directly from the outside of supported processes in the library.

•Interpretability of Complex Ensembles: Despite stacked ensembles' improved prediction precision, there is a trade-off in interpretability. That ensemble models are composed of aggregate natures—most notably those comprising different algorithms—can make tracing specific predictions back to specific features challenging. This becomes a hindrance when decision-makers require explainability for regulation or transparency purposes.

•Time Constraints: Although there have been automated time limits for training models in AutoGluon, in some cases like handling very large or complex data sets, training times remain prohibitive. This calls for careful management of resource usage and model complexity.

These constraints underscore the need for striking a balance between usability and the demand for deep, tailored specifics that some users may need.

7.3 Ethical Considerations

Buyer modeling inevitably brings in some ethical concerns, especially if used in making decisions using personal data:

• Demographic Bias in Data: The majority of datasets, particularly those being collected from web pages, are vulnerable to skewed demographic data in the way of underrepresentation or continuation of past bias. For example, if past promotions were based on certain regions or income levels, the model tends to bias towards these areas and income groups, further embedding present disparities.

• Privacy Concerns: Collection of contextual and behavior data raises high privacy risks, especially with deeper tracking of behavior. While anonymization minimizes privacy risks, it does nothing to eliminate fully concerns about consent of users, use of the data, and surveillance capability. Strong ethical expectations and good practices in data management are needed to ensure appropriate application of this information.

•Fairness and Transparency: There is more need for fairness and transparency in the prediction models. As big business decisions (personalized advertising, customer profiling, etc.) are made using AutoGluon and other such systems, detecting unintended discriminatory results and ensuring that decisions become explainable in a transparent manner to the stakeholders is absolutely vital.

In order to address such concerns, one must implement such protections as fairness audits, anonymization processes, and transparency reporting, combined with adherence to the respective legal frameworks (e.g., GDPR, CCPA).

Future Work

8.1 Integration of Real-Time Contextual Information

With growing personalized marketing and e-commerce, the integration of real-time contextual information such as streaming data from websites or mobile applications could radically enhance the process of buyer modeling. Future research should explore:

•Real-Time Session Information: Blending live user behavior such as clicks, cart additions, and browsing history to provide real-time predictions for tailored offers and recommendations.

• Event-Driven Pipelines: Leveraging streaming technologies (e.g., Apache Kafka, AWS Kinesis) to pipe real-time contextual data into the model and allow for dynamic refresh of predictions and model parameters.

• Adaptive Models: Building models that update in real time based on changing user behavior, thus ensuring recommendations and targeting approaches always reflect the newest information.

This would also increase the personalization of user experiences and drive more engagement and conversion.

8.2 Integration with Deep Learning Models for Image/Text Signals

a second path to future growth involves integrating deep learning models that handle unstructured data such as images or text signals. This will extend the realm of buyer modeling beyond structured tabular data:

• Image Signals: Deep learning algorithms (e.g., CNNs) can be used to analyze product images or customer-uploaded images to provide improved product suggestions based on visual attributes (e.g., color, style, or product category).

•Indicators of Text: Architectures such as transformers or BERT-based embeddings could analyze textual content including customer feedback, search queries, or product descriptions to enhance feature extraction. This would enable a more precise interpretation of buyer preference, emotion, and intent.

With the union of tabular data and deep learning features, we can construct richer and more intricate buyer profiles, which translate into more focused marketing and more personalized experiences.

8.3 Comparison to Other AutoML Pipelines or Hand-Crafted Pipelines

Even though AutoGluon has been shown to achieve high performance on most buyer modeling tasks, future research should compare its performance to other popular AutoML pipelines such as:

•H2O.ai: Characterized by its scalability and domain expert models in areas such as time series forecasting.

•TPOT: An AutoML library based on a genetic algorithm with more customization flexibility, especially in evolutionary model search.

DataRobot: An enterprise-level AutoML tool with strong enterprise support and model explainability features.

Comparing AutoGluon to such tools, future work can infer relative strengths and weaknesses and offer insights into which tools are best suited to which types of buyer modeling tasks.

Also, exploring bespoke AutoML pipelines (e.g., with bespoke preprocessing or model choice) could help create a more accurate sense of where automated systems can do better than normal, hand-crafted solutions. This will allow more flexible, domain-aware configurations, especially for complex tasks or data sets.

Conclusion

In our study, we explored the use of demographic, behavioral, and contextual data integration for end-to-end buyer modeling using AutoGluon, an automatic machine learning framework. Integration of such diverse sources of data offers a strong approach of action for firms seeking to better understand and predict customer behavior. With the help of AutoGluon, organizations are exposed to the following benefits:

Recap of Benefits of Multi-Source Integration Using AutoGluon

• Deep Data Fusion: The integration of dissimilar data types from demographics to browsing history, to real-time context, makes it possible for a 360-degree perspective on the buyer. A large set of data, therefore, gives rise to better predictions and underpins the ability to uncover subtle relationships between customer attributes and behavior.

• Efficiency Automation: AutoGluon's pipeline automation eliminates time and labor spent on time-consuming manual feature engineering or complex hyperparameter optimization. It decreases the effort of modeling, conserving resources and time with little performance loss.

• Scalability: AutoGluon scales well with large datasets, both supporting small as well as massive datasets without problems, allowing companies of all sizes to adopt state-of-the-art machine learning techniques without enormous computational loads.

Holistic Buy-side Insight Equals Enhanced Personalization, Retention, and Targeting

Companies are able to take full advantage of several streams of data brought together under one predictive framework:

•Tailor Customer Engagements: With a complete picture of buying behavior, there is more effective matching of product, marketing campaigns, and targeted promotion to individual buyer needs.

•Increase Retention: Behavior pattern knowledge enables companies to identify likely churn risks and proactively engage with customers through tailored retention initiatives.

•Improved Targeting: Customer segmentation by different parameters (e.g., demographics, behavior, time) improves the focus of marketing campaigns, making them more efficient and effective. It leads to efficient use of marketing budgets and improved customer acquisition outcomes.

Promotion of Low-Code AI/ML Solutions for Strong Modeling Adoption by Companies

Application of low-code AI/ML like AutoGluon assists organizations in tapping machine learning's actual potential without the investment of much data science expertise. Such libraries allow teams to:

•Use Powerful Models: Leverage heavyweight data-driven models with minimal technical bandwidth, thereby having the team focus on strategic ramifications rather than technology complexity.

•Rapid Response to Change in Data: Because customer behavior is always changing, AutoGluon self-refreshing updates ensure models stay relevant with a dynamic response to changing business scenarios.

•Drive Data-Driven Decision-Making: Machine learning model quick testing and deployment makes high-end analytics adoption accessible for everyone, enabling business functions through data-driven decision-making culture.

In short, adding multi-source information and utilizing machine learning automation technology like AutoGluon is a revolutionary move for businesses willing to advance their personalization, retention, and targeting efforts. By adopting this kind of tool, businesses are able to create new avenues towards growth, innovation, and customer satisfaction in today's fast data-driven environment.

REFERENCES:

1. Shoaib, H. A., Rahman, M. A., Maua, J., Rahman, A., Mridha, M. F., Kim, P., & Shin, J. (2025). An Enhanced Deep Learning Approach to Potential Purchaser Prediction: AutoGluon Ensembles for Cross-Industry Profit Maximization. IEEE Open Journal of the Computer Society.
2. Zhang, Y., Yu, Q., & Zhang, L. (2024). User Purchase Intention Prediction Based on Improved Deep Forest. CMES-Computer Modeling in Engineering & Sciences, 139(1).
3. Lu, J., Long, Y., Li, X., Shen, Y., & Wang, X. (2024, September). Hybrid model integration of lightgbm, deepfm, and din for enhanced purchase prediction on the elo dataset. In 2024 IEEE 7th International Conference on Information Systems and Computer Aided Education (ICISCAE) (pp. 16-20). IEEE.
4. Shoaib, Hashibul Ahsan, et al. "An Enhanced Deep Learning Approach to Potential Purchaser Prediction: AutoGluon Ensembles for Cross-Industry Profit Maximization." IEEE Open Journal of the Computer Society (2025).
5. Shoaib, Hashibul Ahsan, Md Anisur Rahman, Jannatul Maua, Ashifur Rahman, M. F. Mridha, Pankoo Kim, and Jungpil Shin. "An Enhanced Deep Learning Approach to Potential Purchaser Prediction: AutoGluon Ensembles for Cross-Industry Profit Maximization." IEEE Open Journal of the Computer Society (2025).
6. Li, Z., Xie, H., Xu, G., Li, Q., Leng, M., & Zhou, C. (2021). Towards purchase prediction: A transaction-based setting and a graph-based method leveraging price information. Pattern Recognition, 113, 107824.
7. Shoaib, Hashibul Ahsan, Md Anisur Rahman, Jannatul Maua, Ashifur Rahman, M. F. Mridha, Pankoo Kim, and Jungpil Shin. "An Enhanced Deep Learning Approach to Potential Purchaser Prediction: AutoGluon Ensembles for Cross-Industry Profit Maximization." IEEE Open Journal of the Computer Society (2025)