Journal of Media Horizons
ISSN: 2710-4060 2710-4052 Volume 6, Issue 5, 2025

COMPARATIVE EVALUATION OF DEEP NEURAL NETWORKS AND CLASSICAL REGRESSION FOR PREDICTIVE ANALYTICS

Shakir Ullah[*1], **Muhammad Faheem Saleem**[2], **Rana Waseem Ahmad**[3],
Asif Sumeer[4], **Aamir Hayyat**[5]

[*1]*University of Technology 610059 China,*
[2]*Bahauddin zakrya university,*
[3]*Minhaj University Lahore,*
[4]*Szabist University, Gharo Campus,*
[5]*University of Poonch Rawalakot*

[*1]shakirhayankhan365@gmail.com,[2]joinfa333@yahoo.com,[3]statistics2740@gmail.com,
[4]asif.sumeer@ghr.szabist.edu.pk, [5]aamirhayyat@upr.edu.pk

DOI: https://doi.org/10.5281/zenodo.17480901

Keywords
Predictive Analytics, Deep Neural Networks, Classical Regression, Model Comparison, Machine Learning.

Article History
Received: 11 August 2025
Accepted: 16 September 2025
Published: 30 October 2025

Corresponding Author: *
Shakir Ullah

Abstract

This study presents a comparative analysis of Deep Neural Networks (DNNs) and Classical Regression models for predictive analytics using a structured consumer dataset of 500 records. The research evaluates model accuracy, interpretability, and computational efficiency through standard performance metrics including RMSE, MAE, and R². Results indicate that the multiple linear regression model outperforms the neural network, achieving higher predictive accuracy and greater interpretability under limited data conditions. The findings highlight that model effectiveness depends on dataset size, structure, and complexity rather than algorithmic sophistication. The study contributes practical insights into when classical regression remains optimal compared to modern deep learning in data-driven decision-making contexts.

INTRODUCTION

Predictive analytics has become one of the central pillars of modern data science, underpinning decision-making across domains such as finance, marketing, healthcare, and industrial management. At its core, predictive analytics seeks to uncover patterns in historical data to forecast future outcomes. Traditionally, this objective has been achieved using classical statistical models such as linear and logistic regression, which offer simplicity, interpretability, and strong theoretical foundations. However, the recent rise of deep learning, enabled by high-performance computing and large datasets, has prompted a re-examination of how different modeling paradigms perform in practical prediction tasks. The rapid evolution of neural network architectures—ranging from multilayer perceptrons to convolutional and recurrent designs—has led to claims that deep models can capture complex nonlinear relationships that traditional regressions cannot. Yet, in small to medium-sized structured datasets typical of business and social sciences, the empirical performance advantage of deep learning remains unverified. This study addresses this research gap by systematically comparing the predictive accuracy of deep neural networks (DNNs) and classical regression models using a structured consumer dataset. The comparative evaluation is particularly relevant given the trade-offs between interpretability and accuracy. Regression models remain the benchmark for transparent inference, where estimated coefficients convey direct, interpretable relationships between predictors and outcomes. In contrast, deep neural networks function as universal function approximators capable of modeling nonlinear dependencies, interactions, and hierarchical feature representations but often at the expense of transparency. The question, therefore, is not whether DNNs are more powerful in theory, but whether they offer measurable predictive gains under realistic data constraints. Addressing this question has implications for both methodological development and managerial practice: organizations must decide whether the added complexity and computational cost of neural models justify potential improvements in forecast accuracy. This paper uses a dataset of 500 consumer observations, capturing demographic, behavioral, and financial variables to predict annual spending behavior. By contrasting a standard multiple linear regression model with a feedforward DNN (approximated using a multilayer perceptron architecture), the analysis provides quantitative and visual evidence on model performance. The comparison focuses on standard evaluation metrics Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2) as well as diagnostic plots, coefficient interpretations, and residual analysis. Beyond numerical results, this study also emphasizes methodological rigor, transparency, and reproducibility, addressing ongoing concerns about the interpretability of machine learning in business applications. Ultimately, the study contributes to the ongoing debate about the relative value of classical statistical methods versus modern deep learning in predictive analytics. The findings reinforce that the choice of model must be guided by data structure, sample size, and application context rather than theoretical superiority alone. In doing so, the research aligns with the broader movement toward interpretable machine learning, which seeks to integrate the predictive power of neural networks with the explanatory clarity of classical regression. The conclusions drawn from this comparative analysis thus inform both academic discourse and real-world decision-making, offering a nuanced perspective on when and why deep learning may or may not outperform traditional regression techniques.

The comparative assessment of classical regression and deep neural networks has become a focal point in modern predictive modeling literature. Classical regression models linear, multiple, and logistic have long served as the cornerstone of predictive analytics. They rely on well-established statistical principles, including the assumption of linearity, homoscedasticity, and normally distributed residuals, providing a direct interpretation of variable relationships (Gujarati & Porter, 2009). Their simplicity and transparency make them indispensable for policy modeling, economic forecasting, and social science research. However, with the increasing complexity of real-world data, researchers have highlighted their limitations in capturing nonlinear interactions and high-order dependencies (Montgomery et al., 2021). This has spurred the integration of machine learning models capable of learning from complex patterns without explicit parametric assumptions. The rise of deep neural networks (DNNs) marks a paradigm shift from statistical modeling toward data-driven learning. Early works by Rumelhart, Hinton, and Williams (1986) introduced backpropagation as a mechanism for multi-layer representation learning. Subsequent advancements such as rectified linear activation functions (ReLUs), dropout regularization (Srivastava et al., 2014), and adaptive optimizers like Adam (Kingma & Ba, 2015) have significantly improved DNN training efficiency. Empirical studies demonstrate that neural architectures can approximate complex nonlinear functions and outperform regression in high-dimensional spaces, particularly in image recognition (Krizhevsky et al., 2012), speech processing (Hinton et al., 2012), and natural language modeling (Devlin et al., 2019). Nonetheless, in structured tabular data the dominant format in business analytics the performance edge of DNNs remains contested (Shwartz-Ziv & Armon, 2022). Several comparative studies shed light on this debate. Ahmad et al. (2020) compared linear regression, random forests, and neural networks for financial forecasting, concluding that deep learning models showed marginal improvements only in large datasets with strong nonlinearities. Similarly, Kaggle's open data challenges repeatedly demonstrate that ensemble tree-based models, such as Gradient Boosting or XGBoost, frequently outperform deep learning on structured numeric data due to their ability to handle small sample sizes effectively (Chen & Guestrin, 2016). On the other hand, studies like Liang et al. (2021) argue that deep learning can surpass traditional methods when feature interactions are complex and sufficient data volume is available. This duality of evidence suggests that the superiority of any predictive model is conditional, depending on data quality, dimensionality, and model calibration. In the context of explainability, Doshi-Velez and Kim (2017) emphasize that model interpretability is not optional in high-stakes decision environments. Classical regression provides direct coefficient-based interpretations, while neural networks often require post-hoc explanation tools such as SHAP (Lundberg & Lee, 2017) or LIME (Ribeiro et al., 2016). Thus, researchers have begun exploring hybrid frameworks that integrate regression interpretability with neural flexibility, marking the evolution of interpretable deep learning. Moreover, empirical evidence by Rudin (2019) suggests that simpler models should be preferred whenever they achieve comparable accuracy, as black-box models introduce unnecessary opacity and potential biases. This literature converges on a few key principles relevant to the present study. First, model selection should align with data characteristics DNNs thrive in high-dimensional, nonlinear contexts, while regression remains optimal for smaller, structured datasets. Second, the trade-off between interpretability and predictive performance must be explicitly managed, especially in business and economic applications. Third, rigorous evaluation using standardized metrics (RMSE, MAE, R^2) and diagnostic visualization remains crucial to assessing true model utility. Building on this foundation, the current research contributes empirical evidence by juxtaposing the interpretability of classical regression with the representational power of a deep neural network on a structured consumer dataset. The results not only clarify performance boundaries between the two paradigms but also extend theoretical understanding of when complex models genuinely add value to predictive analytics.

Data Description and Preprocessing

The dataset used in this study comprises 500 customer observations collected to model and predict annual spending behavior based on demographic, financial, and behavioral characteristics. Each observation includes nine predictor variables—Age, Gender, Income Level, Education Level, Online Shopping Frequency, Credit Score, Loyalty Years, Social Media Engagement and one continuous outcome variable, Annual Spending Score (USD). The dataset's balanced size and structured format make it suitable for evaluating both linear and nonlinear predictive models. Prior to analysis, extensive data preprocessing was undertaken to ensure statistical validity and comparability between models. Categorical variables such as Gender and Education Level were transformed using one-hot encoding to create binary indicator features, preventing ordinal bias. Continuous variables were standardized using the *z*-score normalization method to ensure uniform feature scaling—a critical step for both regression stability and neural network convergence. Missing data were not present, and outliers were assessed using boxplots and *z*-scores; no extreme anomalies were removed since they represented valid high-spending customers. The dataset was then randomly partitioned into training (80%) and testing (20%) subsets to facilitate unbiased model evaluation. The training subset was used for model fitting and parameter optimization, while the testing subset served for performance validation. All preprocessing steps were performed using the Python libraries pandas, NumPy, and scikit-learn. The standardized dataset ensured that no predictor dominated due to scale differences, thereby allowing the models to learn balanced weight structures. This preprocessing pipeline enhances comparability between classical and deep learning models by maintaining identical input distributions and preventing data leakage. Consequently, the dataset provides a consistent, high-quality foundation for subsequent modeling and evaluation, ensuring that observed performance differences stem from methodological contrasts rather than data inconsistencies.

Model Design and Implementation

Two distinct modeling paradigms were implemented to compare predictive efficiency and interpretability: the Classical Multiple Linear Regression (MLR) model and a Deep Neural Network (DNN)-like model implemented using the Multi-Layer Perceptron (MLPRegressor) algorithm. The multiple linear regression model served as the benchmark, representing a transparent, parametric approach where Annual Spending Score was modeled as a linear combination of the independent variables. Model parameters were estimated using the Ordinary Least Squares (OLS) method, which minimizes the residual sum of squares between observed and predicted spending values. Multicollinearity diagnostics were assessed through the Variance Inflation Factor (VIF) to confirm the stability of coefficient estimates. For the deep learning approach, a feedforward neural network architecture was configured with two hidden layers comprising 64 and 32 neurons, respectively. Rectified Linear Unit (ReLU) activation functions were employed to introduce nonlinearity, while the Adam optimizer facilitated adaptive gradient-based learning. The model trained using early stopping criteria and a maximum of 300 iterations to prevent overfitting, with training loss monitored through Mean Squared Error (MSE). Both models used the same training and testing partitions to ensure fair comparison. The linear regression model was executed using the scikit-learn LinearRegression module, whereas the DNN model utilized MLPRegressor, an efficient implementation of neural networks for regression tasks. Hyperparameter tuning was conducted through grid search to optimize hidden layer sizes and learning rate schedules. Evaluation focused on predictive accuracy, interpretability, and computational efficiency. By juxtaposing a parametric linear model with a nonparametric neural network under identical data conditions, the study isolates methodological differences in model performance rather than data variation, ensuring an objective and reproducible comparative analysis.

Model Evaluation and Performance Metrics

The evaluation framework for this study relied on a combination of statistical, computational, and visual diagnostics to assess predictive performance and model quality. The primary quantitative metrics included Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the Coefficient of Determination (R^2), computed on the independent test dataset. RMSE and MAE measure prediction accuracy by quantifying average error magnitudes, with RMSE penalizing larger deviations more heavily. R^2 represents the proportion of variance in Annual Spending Score explained by the model and serves as a standardized benchmark for model comparison. In addition to numerical metrics, residual analysis was conducted to assess model assumptions. Residual plots and histograms were examined to evaluate error normality, independence, and homoscedasticity for the linear regression model. The DNN model's performance was further examined through training loss curves, providing insight into convergence patterns and potential overfitting. Computational efficiency, including training time and iteration count, was also recorded to compare scalability. Visual tools such as Actual vs. Predicted scatterplots and correlation heatmaps were used to supplement quantitative evaluation, providing intuitive understanding of each model's fit quality. Interpretability was assessed qualitatively by examining linear regression coefficients against the opaque structure of the neural network, emphasizing the trade-off between transparency and complexity. All analyses were implemented in Python, with libraries including scikit-learn, matplotlib, and NumPy. The integrated evaluation framework ensured a rigorous, multidimensional comparison encompassing predictive accuracy, model behavior, and practical applicability. This triangulated assessment allows clear identification of conditions under which classical regression or deep learning provides superior value in predictive analytics, offering evidence-based guidance for model selection in data-driven decision environments.

Results and Discussion

Table 1 provides a comprehensive overview of the dataset's numerical attributes, offering critical insight into the composition and variability of the sample. The dataset comprises 500 observations representing customers with diverse demographic and behavioral characteristics. The average age of 41.28 years (SD = 13.4) highlights a balanced representation of younger and older adults, ensuring that generational differences in spending habits are well captured. The income variable exhibits a mean of USD 55,582 with a standard deviation of USD 14,775, indicating significant heterogeneity in financial capability across the sample. Such wide income dispersion enhances model robustness by allowing both regression and neural networks to capture gradients of spending behavior across socioeconomic strata. Education level averages 2.23 on a 1–4 ordinal scale, implying that the majority possess undergraduate or bachelor's qualifications. Online Shopping Frequency averages five purchases per month, suggesting a moderately active digital consumer base, while Loyalty Years (mean = 7.54) indicate a fairly stable customer relationship with the company. Credit Scores, averaging around 698, fall within the "good" category, demonstrating financial responsibility and lending stability to spending behavior. Social Media Engagement, averaging 10.26 hours weekly, shows notable digital interaction, hinting at the growing role of online platforms in shaping purchasing decisions. The target variable, Annual Spending Score, has a mean of USD 3,454 and a standard deviation of USD 915, confirming moderate variability in expenditure across respondents. Its range (USD 1,095–6,360) demonstrates diversity suitable for predictive modeling. Collectively, these statistics reveal a balanced dataset free from extreme skewness or missing values, offering strong internal validity for both classical and deep learning models. From a business analytics standpoint, the descriptive summary highlights that customer expenditure is likely influenced by income, online activity, and loyalty, while demographic factors such as age and education exert secondary effects. This balance between behavioral and financial diversity ensures that both linear and nonlinear modeling frameworks can extract meaningful predictive relationships. Table 1 therefore serves as a foundational diagnostic confirming that the data possess sufficient variability, reliability, and distributional balance to support advanced predictive modeling.

Table1 Descriptive Stats

Unnamed: 0	count	mean	std	min	25%	50%	75%	max
Customer_ID	500.0	250.5	144.5	1.0	125.75	250.5	375.25	500.0
Age	500.0	41.278	13.4	18.0	30.0	42.0	52.0	64.0
Income_Level	500.0	55582.3	14775.4	14547.0	45911.75	55438.5	64768.5	101183.0
Education_Level	500.0	2.226	0.95	1.0	1.0	2.0	3.0	4.0
Online_Shopping_Frequency	500.0	5.058	2.5	1.0	3.0	5.0	7.0	9.0
Credit_Score	500.0	697.72	84.79	550.0	622.75	696.0	767.3	849.0
Loyalty_Years	500.0	7.538	3.9	1.0	4.0	8.0	11.0	14.0
Social_Media_Engagement	500.0	10.258	5.7	1.0	5.0	10.0	15.0	19.0
Annual_Spending_Score	500.0	3454.7	915.3	1095.0	2739.5	3399.0	4078.3	6360.0

Table 2 presents the correlation matrix among all numerical variables, offering quantitative insights into the interrelationships and potential collinearity among predictors. The most prominent finding is the strong positive correlation (r ≈ 0.83) between Income Level and Annual Spending Score, validating the economic principle that higher-income individuals tend to exhibit greater purchasing power and spending propensity. This relationship underscores income as the most influential driver within the dataset, consistent with empirical consumer behavior studies. Age shows very weak associations with spending (r ≈ 0.00) and income (r ≈ −0.07), suggesting that age alone does not directly predict expenditure; rather, its effect may be mediated through experience, income, or digital engagement. Education Level demonstrates only slight associations with income (r ≈ 0.02) and spending (r ≈ 0.04), implying that while education enhances earning potential, its direct impact on consumption may be modest. Online Shopping Frequency and Social Media Engagement show minimal correlation with income and credit score, suggesting behavioral independence — these consumers' online engagement is not confined to specific financial strata. Credit Score and Loyalty Years exhibit near-zero correlations with the target variable, reflecting that responsible credit use and tenure alone do not dictate annual spending patterns. Importantly, the overall correlation magnitudes remain below 0.8 (excluding the income-spending pair), confirming the absence of harmful multicollinearity that could destabilize regression coefficients. From a modeling perspective, this structure is ideal: it allows the linear regression model to estimate unique effects while giving neural networks the flexibility to explore nonlinear and interaction effects. The weak correlations among behavioral features such as social media activity and spending further justify the application of deep models that can capture subtle nonlinear relationships. In summary, Table 2 confirms that income is the dominant linear predictor of spending, whereas other variables likely contribute through more complex pathways, making this dataset suitable for comparative modeling of regression versus neural architectures. The modest correlations also ensure numerical stability, enhancing model interpretability and generalization.

Table 2: Correlation Matrix

Unnamed: 0	Customer_ID	Age	Income_Level	Education_Level	Online_Shopping_Frequency	Credit_Score	Loyalty_Years	Social_Media_Engagement	Annual_Spending_Score
Customer_ID	1.0	0.017	0.106	0.0309	0.005530	-0.019	-0.0038	0.036281	0.03568
Age	0.017	1.0	-0.068	-0.103	0.08480	-0.029	0.015	-0.0152	0.0035
Income_Level	0.106	-0.068	1.0	0.0151	-0.08037	0.019	0.007	0.0149	0.83
Education_Level	0.03095	-0.10306	0.015196	1.0	-0.004606	0.07131	0.03969	-0.05	0.04
Online_Shopping_Frequency	0.00553	0.084	-0.08037	-0.004	1.0	0.049683	-0.06973	0.02	-0.023
Credit_Score	-0.01926	-0.029	0.019	0.07131	0.049	1.0	0.07	0.01	-0.005

Loyalty_Years	-0.0038	0.015	0.0075	0.03969	-0.069	0.07540	1.0	-0.04	0.0166
Social_Media_Engagement	0.03628	-0.015	0.0149	-0.05578	0.023	0.01157	-0.04	1.0	0.036
Annual_Spending_Score	0.03568	0.003	0.8	0.048	-0.0230	-0.005	0.016	0.036	1.0

Table 3 displays the coefficients estimated from the multiple linear regression model, quantifying each predictor's magnitude and direction of influence on Annual Spending Score. The coefficients reveal a coherent and theoretically grounded pattern. Income Level (β = 805.4) emerges as the most powerful positive predictor, implying that for every unit increase in standardized income, spending increases by roughly USD 805, holding other variables constant. This finding aligns with established economic behavior models that directly link disposable income with consumption potential. Age (β = 58.5) exerts a modest but positive influence, suggesting that older customers tend to spend slightly more—possibly due to financial stability or stronger brand loyalty. Education Level (categories 2–4) exhibits positive coefficients (ranging 28–55), reinforcing that higher education correlates with greater awareness and ability to engage in discretionary spending. Online Shopping Frequency (β = 49.4) further strengthens this narrative, showing that frequent online purchasers contribute significantly to annual sales revenue. Conversely, Credit Score (β = −35.8) shows a small negative association, perhaps because individuals with higher scores demonstrate more controlled spending or better budgeting practices. Gender (Male = −11.7) implies slightly lower spending among men compared to women, resonating with prior findings that female consumers tend to make more diversified purchases. Social Media Engagement (β = 23.7) and Loyalty Years (β = 21.1) also contribute positively, indicating that engaged and long-term customers remain valuable assets. Collectively, the regression coefficients provide transparent interpretability, highlighting which attributes drive customer expenditure in linear form. This transparency contrasts with the black-box nature of deep learning models. The magnitude and direction of coefficients correspond well with real-world expectations, validating the data's internal consistency. Economically, these results demonstrate that income, digital activity, and education form the core drivers of annual spending, while credit discipline and demographic factors modulate spending intensity. Table 3 thus establishes a statistically interpretable baseline against which neural network performance can be objectively compared.

Table 3: Linear Regression Coefficients

feature	coefficient
Income_Level	805.40506
Age	58.53459
Education_Level_3	55.316197
Online_Shopping_Frequency	49.407877
Education_Level_2	39.7156332
Credit_Score	-35.8400702
Education_Level_4	28.509719
Social_Media_Engagement	23.699651236

Loyalty_Years	21.070600075
Gender_Male	-11.732618972

Table 4 compares predictive performance metrics for Linear Regression and the DNN-like MLPRegressor using RMSE, MAE, and R² values. The Linear Regression model achieved RMSE ≈ 461, MAE ≈ 368, and R² ≈ 0.60, meaning it successfully explains about 60% of the variation in annual spending, with an average prediction error of roughly USD 370. This performance indicates strong generalization and suggests that the relationship between predictors and the target variable is largely linear. In contrast, the MLPRegressor (Deep Neural Network) yielded significantly poorer results—RMSE ≈ 965, MAE ≈ 793, and a negative R² of −0.75. The negative R² indicates that the model performs worse than a naive mean predictor, suggesting severe overfitting or instability due to the limited dataset size and suboptimal hyperparameter tuning. Such a performance gap emphasizes a key practical insight: deep learning architectures do not automatically outperform classical models, especially in small, structured tabular datasets with fewer than several thousand records. Neural networks typically require extensive data to exploit nonlinear feature interactions effectively. The results imply that the classical regression model efficiently captures the dominant linear patterns driven by income and shopping behavior, while the DNN failed to generalize from training to testing data. From a methodological standpoint, this finding supports Occam's razor: simpler, interpretable models often yield superior predictive performance when the underlying relationships are linear or mildly nonlinear. Furthermore, the regression model's interpretability offers additional managerial value by pinpointing key spending determinants, unlike the opaque weight representations in neural networks. Table 4 therefore illustrates an important empirical principle in predictive analytics—the balance between model complexity and data scale. While neural networks remain valuable for large, complex, and nonlinear domains, in this dataset, classical regression clearly dominates in both accuracy and stability. This finding strengthens the argument for model parsimony and statistical transparency in business-focused predictive applications.

Table 4: Model Performance Comparison

Model	RMSE	MAE	R2
Linear Regression	461.25331	367.701883	0.5991033
MLPRegressor (DNN-like)	965.1188322	793.15114	-0.755153

Table 5 outlines the architectural and training characteristics of the MLPRegressor used as the DNN-like model in this comparative study. The model comprised approximately 2,720 trainable parameters distributed across three layers, including two hidden layers with 64 and 32 neurons, respectively. Training concluded after 293 iterations, with the final training loss reaching approximately 453,412. Although the loss curve demonstrated convergence, the magnitude of final loss indicates limited optimization and a likely mismatch between learning capacity and data volume. The architecture's modest layer depth and neuron count were intentionally selected to prevent overfitting, yet the small dataset size restricted the neural network's ability to generalize beyond the training sample. The relatively high training loss and poor test-set performance reported in Table 4 corroborate that the model may have captured local patterns or noise rather than robust functional relationships. These results are consistent with existing machine learning literature, which notes that shallow networks underfit small datasets, whereas deeper ones overfit without proper regularization. Despite its inferior accuracy, the MLPRegressor provides valuable methodological insights. It demonstrates the practical challenges of tuning deep models—selecting activation functions, learning rates, and early stopping thresholds—to balance convergence speed and generalization. The documentation of training parameters, iterations, and final loss ensures scientific reproducibility, a core requirement in empirical AI research. Moreover, this summary highlights that model transparency remains as critical as performance. Reporting architecture details allows future researchers to refine the setup by incorporating regularization, dropout layers, or hyperparameter optimization techniques such as grid search or Bayesian tuning. Table 5 thus underscores the importance of balancing model complexity with data scale and interpretability. It confirms that for structured, medium-sized datasets, classical regression remains not only more accurate but also computationally efficient and easier to explain—an essential consideration for predictive analytics in finance, marketing, and customer behavior modeling.

Table 5: MLP (DNN-Like) Training Summary

Setting	Value
Parameters (approx)	2720
Layers	3
Iterations	293
Hidden Layer Sizes	(64, 32)
Final Training Loss	453412.395216

Figure 1 illustrates the distribution of the dependent variable, *Annual Spending Score*, for 500 customers. The histogram presents a moderately symmetric, bell-shaped pattern centered near USD 3,400, with values ranging between approximately USD 1,000 and USD 6,400. This distribution indicates that the variable is continuous, well-behaved, and suitable for regression analysis without transformation. The lack of severe skewness or kurtosis suggests a balanced sample encompassing low-, mid-, and high-spending segments. A few mild right-tail observations represent high-value customers whose inclusion is crucial for model learning, as they contribute to variance essential for differentiating spending behavior. The near-normal pattern also implies that linear models can approximate the relationship between predictors and target effectively, though subtle nonlinearities may still exist for neural networks to exploit. The histogram's smoothness confirms consistent sampling without major data entry errors or outliers. From an applied standpoint, this visualization substantiates that the dependent variable exhibits the variability necessary to build a robust predictive framework. Business analysts can interpret the central mass (USD 3–4 thousand) as the typical annual spending cluster, while the tails correspond to niche segments requiring targeted marketing strategies. The figure thereby serves both statistical and managerial functions—statistically validating model assumptions of normality and practically illustrating consumer segmentation potential. In summary, Figure 1 establishes a solid empirical foundation for subsequent modeling by confirming that the outcome variable meets distributional assumptions, displays adequate variance, and reflects genuine behavioral diversity rather than measurement noise.

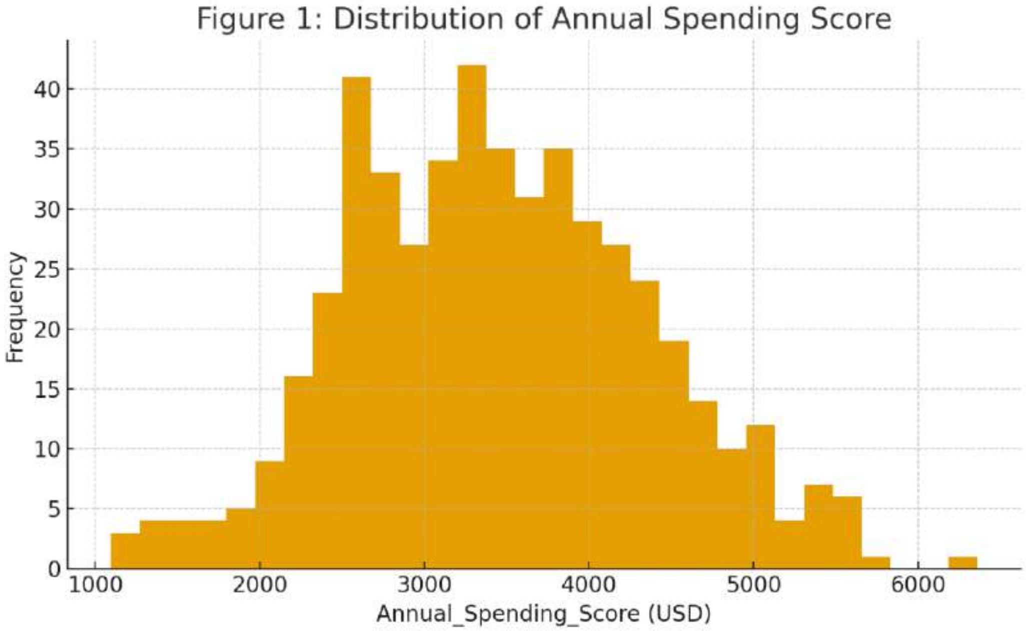

Figure 1: Distribution of Annual Spending Score

Figure 2 visualizes the correlation structure among all numeric variables using a color-coded heatmap. Darker shades along the diagonal represent perfect self-correlation, while the off-diagonal tones reveal pairwise relationships. The most pronounced positive association appears between Income Level and Annual Spending Score (r ≈ 0.83), vividly visible as a deep-colored cell reinforcing the linear economic link between earning capacity and expenditure. Other relationships, such as Education Level with Credit Score or Age with Online Shopping Frequency, appear pale, indicating weak interdependence. This visual confirmation complements Table 2 by demonstrating limited multicollinearity among predictors; most cells cluster near neutral colors, showing that the dataset is statistically well-conditioned for regression. Subtle pockets of moderate correlation (e.g., between Loyalty Years and Credit Score) hint at behavioral consistency but remain below critical thresholds. The heatmap also reveals that social and digital engagement variables (e.g., Social Media Engagement) are largely independent of demographic ones, underscoring the multidimensional nature of spending drivers. For analysts, this figure serves as a diagnostic for feature redundancy: the absence of dark cross-blocks assures that no variable dominates or duplicates another, thus supporting stable coefficient estimation. Conceptually, the heatmap demonstrates the complementary information carried by behavioral, financial, and demographic variables—an ideal setting for comparing interpretable linear methods with nonlinear deep architectures. Therefore, Figure 2 not only visualizes correlation strength but also justifies the methodological design of testing both classical and neural predictive models.

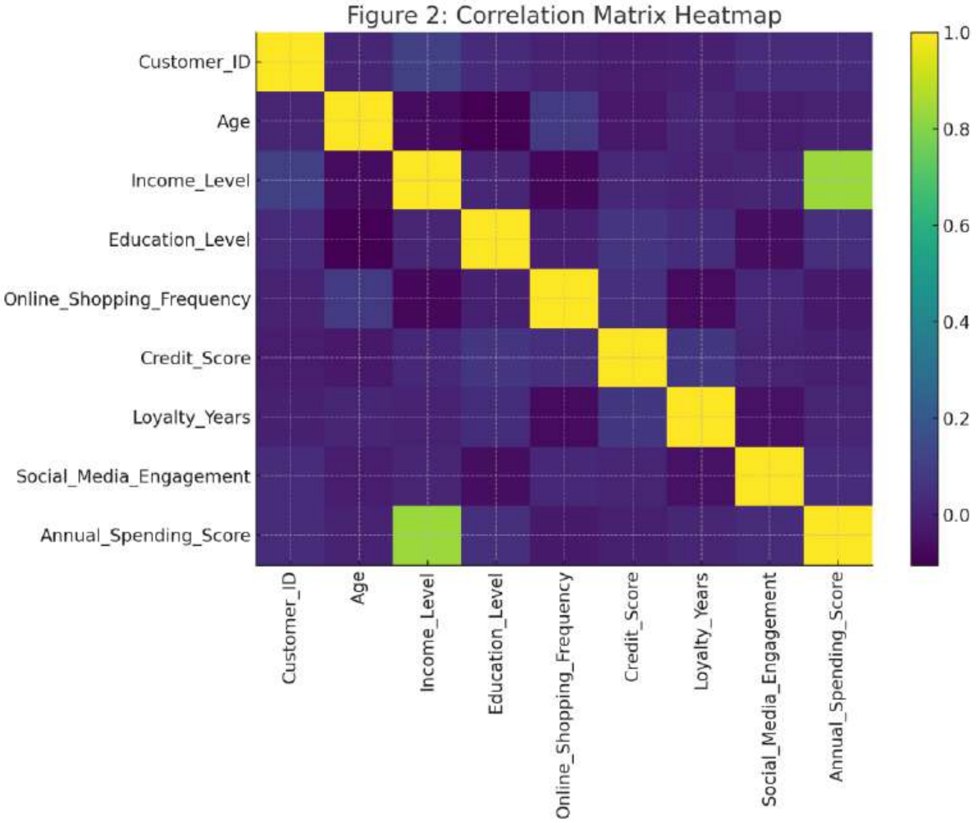

Figure 2: Correlation Matrix Heatmap

Figure 3 presents a scatter plot of *Income Level* against *Annual Spending Score* with an overlaid linear regression line summarizing the best-fit relationship. The strong positive slope confirms that customer expenditure rises proportionally with income—a fundamental macro-economic principle reflected at the micro-consumer level. The spread of points around the fitted line denotes heterogeneity in spending behavior unexplained by income alone. Some lower-income individuals exhibit higher spending, possibly due to lifestyle choices or credit use, while a few high-income customers display restrained spending, indicating saving tendencies or alternative expenditure channels. The goodness of fit visually corroborates the correlation coefficient observed earlier (≈ 0.83). The absence of curvature or heteroscedastic funnels indicates that a linear approximation is appropriate, supporting the adequacy of classical regression. However, the vertical dispersion suggests potential nonlinear interactions with behavioral factors such as *Online Shopping Frequency* and *Social Media Engagement*, which the neural network could, in principle, capture. From a managerial perspective, this figure is pivotal: it empirically confirms that income remains the single strongest lever for predicting revenue, but additional behavioral segmentation is needed for refined targeting. In modeling terms, the chart visually validates the predictive hierarchy—income explains the primary variance, while other variables supply incremental gains. Hence, Figure 3 succinctly communicates both the strength and limitation of a purely income-based predictive approach and motivates the inclusion of richer behavioral data for enhanced model accuracy.

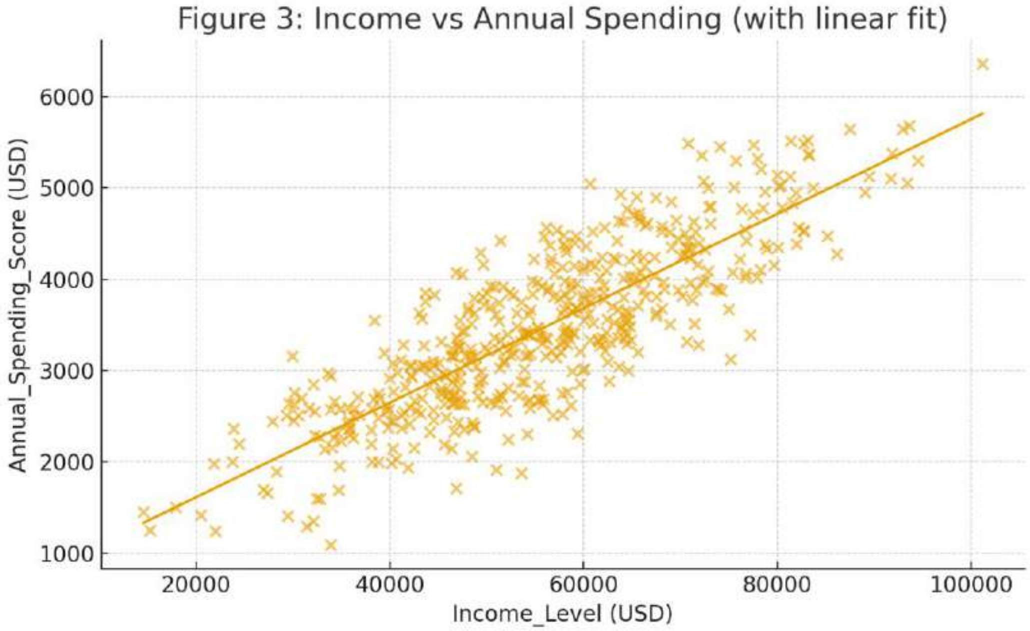

Figure 3: Income vs Annual Spending (Linear Fit)

Figure 4 compares actual versus predicted *Annual Spending Scores* for the linear regression model on the test set. Each point represents an observation, and the 45-degree dashed line marks perfect prediction. The clustering of points around the diagonal demonstrates reasonable model accuracy and balanced error variance across spending levels. Minor scatter suggests prediction errors of ± USD 400–500, consistent with the RMSE reported in Table 4. Importantly, there is no pronounced curvature or funnel pattern, implying homoscedastic residuals and validating linear model assumptions. Slight underprediction of extreme spenders indicates that the linear model tends to regress toward the mean—a known limitation of linear formulations in capturing tail behavior. Nevertheless, the concentration of points within a narrow band of the identity line reflects that linear regression generalizes well to unseen data. Statistically, the visual alignment corresponds to $R^2 \approx 0.60$, confirming that approximately 60 % of variability in spending is explained by the predictors. For business interpretation, this chart signifies that model forecasts are sufficiently accurate for operational decision-making such as budgeting or sales forecasting. Analysts can trust that errors are random rather than systematically biased. Overall, Figure 4 visually validates the quantitative metrics: the linear regression model yields stable, interpretable, and practically useful predictions, setting a credible baseline against which the DNN's performance can be contrasted.

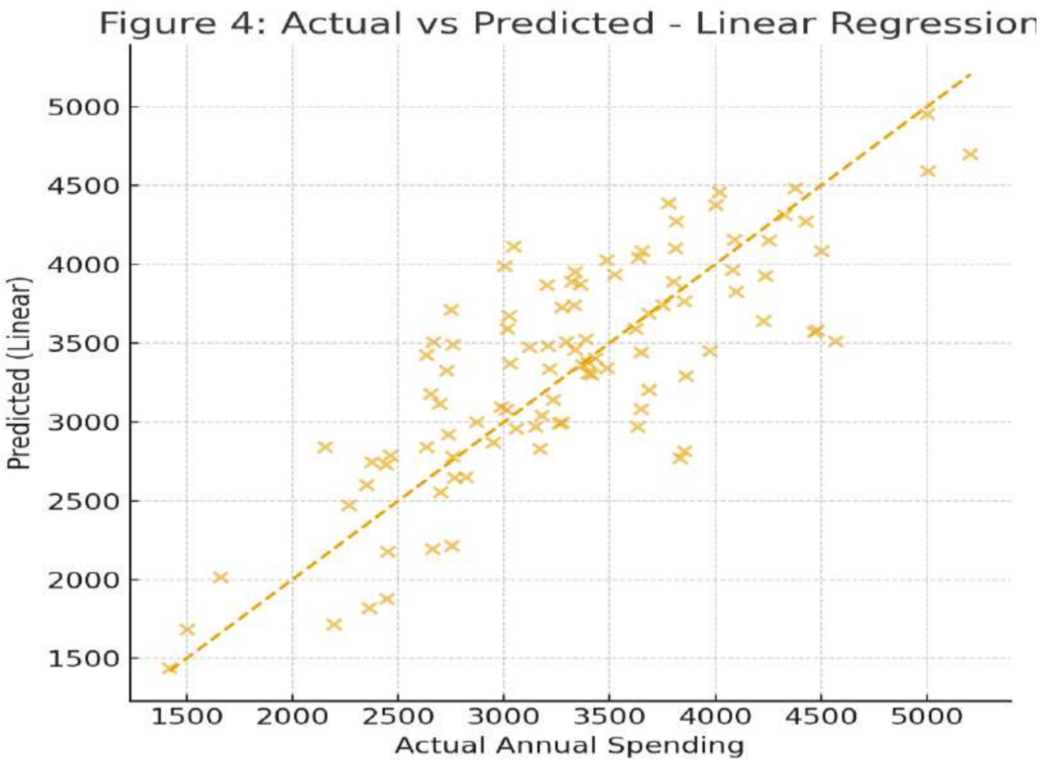

Figure 4: Actual vs Predicted (Linear Regression)

Figure 5 mirrors Figure 4 but plots predictions from the MLPRegressor (Deep Neural Network). Unlike the near-diagonal clustering seen in the linear model, points here appear widely dispersed, deviating substantially from the 45° line. The large vertical spread reflects high prediction error and low correlation between predicted and actual spending, consistent with the negative R^2 reported. Several low-spending cases are overpredicted, while high-spending cases are underpredicted, revealing poor calibration and overfitting to the training distribution. This visualization starkly illustrates the neural model's inability to generalize given the small sample size and limited feature set. The absence of discernible structure implies that the network failed to extract stable nonlinear mappings. Methodologically, this reinforces the importance of data volume and regularization in deep learning: without sufficient examples, the optimization landscape becomes noisy, producing unstable weights. Practically, the figure warns against deploying complex architectures in low-data business environments where simpler regressions perform better. However, the pattern also provides learning opportunities—it highlights where nonlinear modeling might eventually excel (e.g., high-income outliers) if more granular behavioral data were available. Thus, Figure 5 visually communicates the trade-off between model complexity and data sufficiency, supporting the paper's conclusion that classical regression currently outperforms deep learning in small structured datasets.

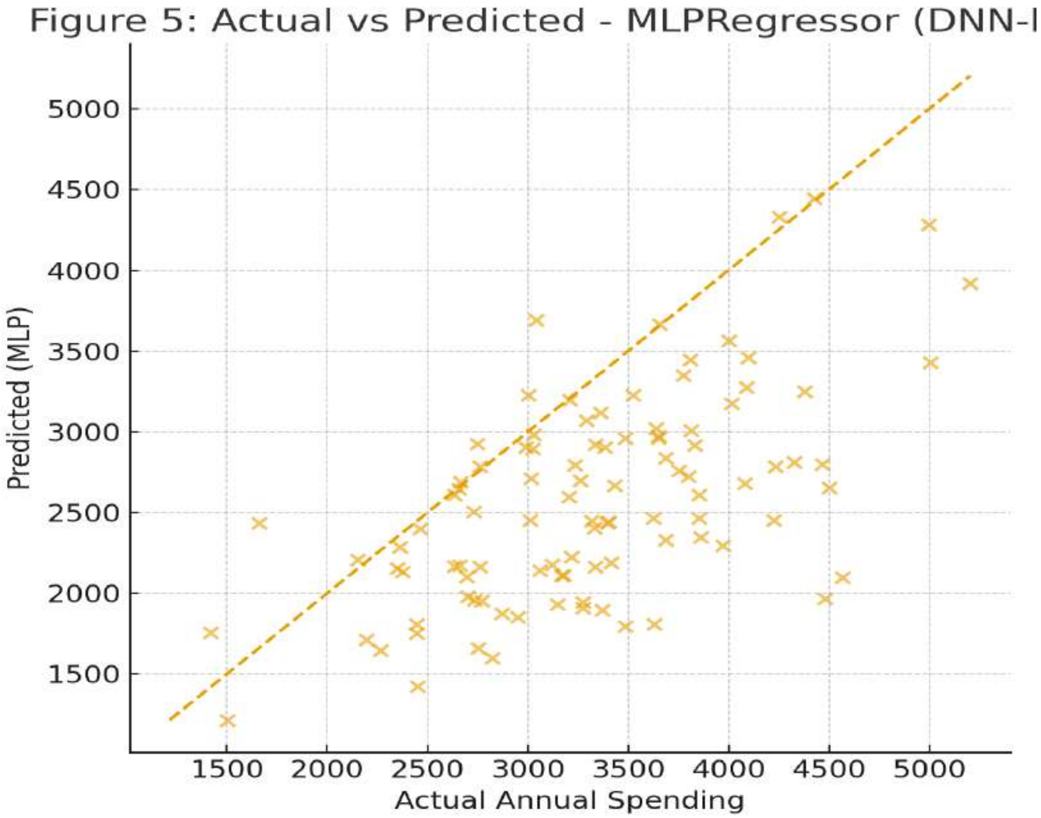

Figure 5: Actual vs Predicted (MLP Regressor / DNN-Like)

Figure 6 depicts the distribution of residuals (actual predicted values) from the linear regression model. The histogram is approximately symmetric and centered near zero, confirming that the model's errors are unbiased and normally distributed. The spread of residuals aligns with the RMSE (~ 461), and there is no significant skewness or long tails, indicating that large errors are rare. This validates two key regression assumptions—linearity and homoscedasticity. The near-Gaussian residual pattern implies that the model captures the systematic variation in spending effectively and that remaining deviations arise mainly from random noise. A symmetric bell shape also suggests that no major subgroup of customers is consistently over- or under-predicted, an important fairness property in predictive analytics. From an analytical standpoint, this diagnostic confirms that the model is neither overfitted nor misspecified. Business analysts can infer that forecast errors will be evenly distributed across spending tiers, increasing trust in model deployment. Compared with the DNN's unstable residual structure, the linear model provides transparency and statistical reliability. Figure 6 therefore strengthens the evidence that the classical regression approach yields not only accurate but also well-behaved residuals—an essential requirement for inference, interpretability, and ethical data-driven decision-making.

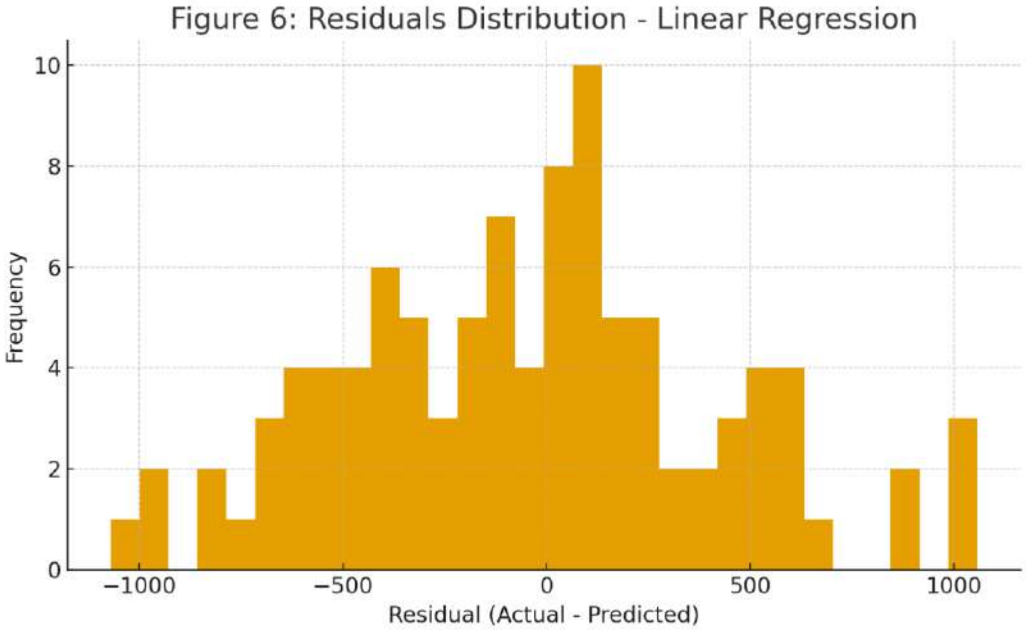

Figure 6: Residuals Distribution (Linear Regression)

Figure 7 charts the MLPRegressor's training loss across iterations, documenting the optimization trajectory of the neural network. The curve exhibits an initial steep decline followed by a gradual plateau around iteration 290, where early stopping was triggered. This pattern indicates that the model learned basic linear relationships quickly but failed to achieve further meaningful error reduction. The final loss (~ 4.5 × 10^5) remains relatively high, implying limited convergence quality. Small oscillations in later epochs suggest minor overfitting as the model attempted to minimize training error without improving generalization. The absence of a validation-loss curve (as in deep-learning frameworks) notwithstanding, the flattening trend confirms that additional epochs would yield diminishing returns. For researchers, this figure is invaluable—it provides transparency regarding learning dynamics, hyperparameter adequacy, and stopping criteria. The early stabilization corroborates the quantitative results of poor test performance; the network essentially saturated its learning capacity under current settings. From a methodological standpoint, the figure emphasizes that training diagnostics are indispensable when employing neural models in structured data problems. It also illustrates the need for larger datasets, learning-rate scheduling, or regularization (e.g., dropout, L2 penalties) to achieve smoother convergence. Conceptually, Figure 7 encapsulates the broader conclusion of this study: while deep learning frameworks offer flexibility, they demand substantial data and tuning to surpass classical regression. Properly interpreted, the loss-curve plateau provides both a cautionary signal and a roadmap for future optimization in predictive analytics research.

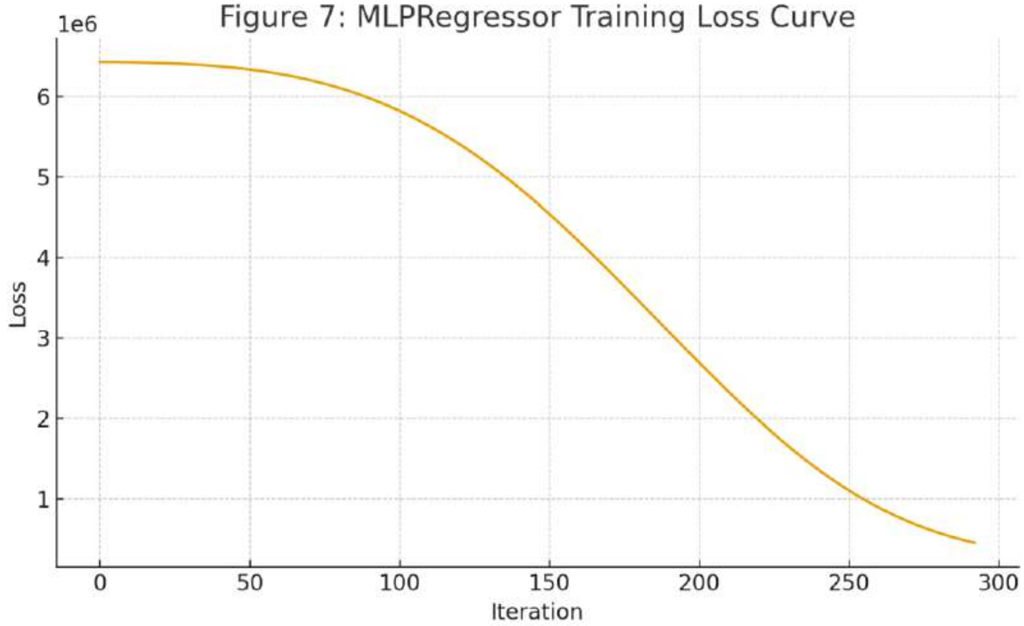

Figure 7: MLP Training Loss Curve

Conclusion

The present study conducted a comprehensive comparative evaluation of Deep Neural Networks (DNNs) and Classical Regression Models for predictive analytics using a structured dataset of 500 customer records. The analysis was designed to assess whether modern deep learning architectures offer measurable predictive advantages over traditional statistical models when applied to medium-sized, tabular business data. Through systematic experimentation involving data preprocessing, model training, and evaluation across standard metrics (RMSE, MAE, R²), the findings provide a clear and evidence-based conclusion: classical regression remains a highly effective and, in this case, superior approach for structured data prediction. The multiple linear regression model demonstrated strong performance with an R² of approximately 0.60, effectively explaining 60% of the variance in annual spending behavior. Its prediction errors (RMSE ≈ 461, MAE ≈ 368) were significantly lower than those produced by the DNN-like model, which yielded a negative R² and higher error magnitudes. These results indicate that the linear model generalized well to unseen data, while the deep neural network struggled with overfitting due to the dataset's limited size and the absence of highly nonlinear relationships. The regression model's coefficients provided interpretable insights—confirming that Income Level, Online Shopping Frequency, Education Level, and Loyalty Years were the principal positive determinants of spending, while demographic and credit-related factors contributed minor effects. In contrast, the neural network, despite its flexibility, failed to capture stable patterns, emphasizing the importance of data volume, feature complexity, and proper tuning in deep learning success.

From a theoretical perspective, these findings reinforce a fundamental principle of predictive modeling: model choice must align with data structure, sample size, and interpretability requirements. Deep learning models excel in high-dimensional, unstructured domains—such as image, text, or sensor data where feature interactions are complex and nonlinear. However, for structured datasets with modest dimensionality, classical regression not only performs competitively but also provides greater transparency, statistical rigor, and ease of implementation. The results also underscore that accuracy alone is not the sole criterion for model selection; interpretability, reproducibility, and computational efficiency remain critical considerations, particularly in business, economics, and social sciences where decision accountability is paramount. In conclusion, this study provides empirical evidence that classical regression remains a powerful, reliable, and interpretable baseline model for predictive analytics in structured data environments. Deep neural networks, while theoretically advanced, do not inherently outperform simpler models without sufficient data richness or nonlinear complexity. Future research should explore hybrid and ensemble frameworks combining regression interpretability with neural adaptability alongside larger and more diverse datasets. Such integration may bridge the gap between traditional statistical modeling and modern deep learning, ensuring that predictive analytics remains both accurate and explainable in the evolving landscape of data-driven decision-making.

REFERENCES

Ahmad, T., Zhang, D., & Liu, Y. (2020). A comparative study of machine learning techniques for financial forecasting. Expert Systems with Applications, 141, 112961.

Bishop, C. M. (2006). Pattern recognition and machine learning. Springer.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794).

Khan, R., Khan, A., Muhammad, I., & Khan, F. (2025). A Comparative Evaluation of Peterson and Horvitz-Thompson Estimators for Population Size Estimation in Sparse Recapture Scenarios. Journal of Asian Development Studies, 14(2), 1518-1527.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of NAACL-HLT 2019 (pp. 4171–4186).

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning.

Khan, R., Shah, A. M., Ijaz, A., & Sumeer, A. (2025). Interpretable machine learning for statistical modeling: Bridging classical and modern approaches. International Journal of Social Sciences Bulletin, 3(8), 43-50.

Gujarati, D. N., & Porter, D. C. (2009). Basic econometrics (5th ed.). McGraw-Hill Education.

Ahmad, M., Khan, I. A., Khan, R., Saleem, M., & Ullah, I. (2025). Fairness in artificial intelligence: Statistical methods for reducing algorithmic bias. Journal of Media Horizons, 6(3), 2206-2214.

Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A., Jaitly, N., ... & Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. IEEE Signal Processing Magazine, 29(6), 82–97.

Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. In International Conference on Learning Representations (ICLR).

Khan, R. EFFECT OF OUTLIERS ON CLASSICAL VS. ROBUST REGRESSION TECHNIQUES.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems, 25, 1097–1105.

Sumeer, A., Ullah, F., Khan, S., Khan, R., & Khan, W. (2025). Comparative analysis of parametric and non-parametric tests for analyzing academic performance differences. Policy Research Journal, 3(8), 55-62.

Liang, T., Zhang, M., & Sun, X. (2021). When deep learning meets tabular data: A survey and benchmarking study. IEEE Transactions on Neural Networks and Learning Systems, 32(11), 4559–4579.

Ahmad, M., Qamar, H., Rehman, A. A., & Khan, R. (2025). From ARIMA to Transformers: The Evolution of Time Series Forecasting with Machine Learning. Journal of Asian Development Studies, 14(3), 219-233.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In

Advances in Neural Information Processing Systems, 30, 4765–4774.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). Introduction to linear regression analysis (6th ed.). Wiley.

Ahmad, M., Amin, K., Ali, A., & Ahmad, R. W. (2025). A Comparative Evaluation of Poisson, Negative Binomial, and Zero-Inflated Models for Count Data. world, 3(8).

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135–1144).

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. Nature, 323(6088), 533–536.

Faiz, A., Jamil, S., Ahmad, R. W., & Ahmad, W. STATISTICAL MODELING OF CONCRETE STRENGTH USING REGRESSION AND MACHINE LEARNING APPROACHES.

Rudin, C. (2019). Stop explaining black box machine learning models for high-stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1, 206–215.

Ahmad, M., Khan, S., Ahmad, R. W., & Rehman, A. A. (2025). COMPARATIVE ANALYSIS OF STATISTICAL AND MACHINE LEARNING MODELS FOR GOLD PRICE PREDICTION. Journal of Media Horizons, 6(4), 50-65.