

ML1 Telco Customer Analysis - Neural Network & Poisson GLM Documentation

Merun Gugelmann, Khan Gulraiz and Roger Bavibidila

2025-06-02

Contents

1	Introduction	2
2	Background	2
2.1	Research Questions and Hypothesis	2
3	Method	2
3.1	Data Set	2
3.2	Research Process	2
3.3	SVM and LM	3
3.3.1	Linear Model for CLTV	3
3.3.2	SVM for Churn Classification	6
3.3.3	Final Model Training and Evaluation	7
3.3.4	SVM Results	8
3.3.5	Practical Implications	8
3.4	Neural Network Analysis	8
3.4.1	Approach and Methodology	8
3.4.2	Feature Selection Strategy	9
3.4.3	Model Architecture Design	9
3.4.4	Evaluation Metrics and Performance Assessment	9
3.5	Poisson GLM Analysis	10
3.5.1	Approach and Methodology	10
3.5.2	Feature Selection for Count Modeling	10
3.5.3	Model Architecture and Statistical Framework	11
3.5.4	Evaluation Metrics for Count Prediction	11
3.6	General Additive Model & Binomial Churn	11

4 Results	11
4.1 Results Analysis and Business Implications - Neural Networks	11
4.2 Results Analysis and Strategic Insights - Poisson GLM	12
4.3 Comparative Model Insights	12

Abstract Here is the abstract

1 Introduction

2 Background

2.1 Research Questions and Hypothesis

3 Method

3.1 Data Set

The dataset consists of customer data including demographic information (e.g., Gender, Age), service details (e.g., Contract.Type, Internet.Type), and customer behavior (e.g., Tenure.in.Months, Monthly.Charge, Total.Refunds). The response variables we worked with were:

- **CLTV (Customer Lifetime Value):** a continuous variable indicating the lifetime value of a customer.
- **Churn:** a binary classification target (Yes/No) representing whether the customer churned. Missing values were handled by filtering out incomplete records for relevant models. Correlated features were examined using a correlation matrix, and some features were removed to avoid multicollinearity.

3.2 Research Process

since the whole work was divided into 3 parts in order to make that all of us contribute equally to the project. To this end, following research steps were taken for LM and SVM Part. 1. **Regression Modeling for CLTV Prediction:** - Initial linear model using all available predictors. - Variance Inflation Factor (VIF) analysis to assess multicollinearity. - Polynomial transformation (degree 4) on **Tenure.in.Months** to capture non-linear effects. - Stepwise selection based on AIC to identify the most parsimonious model. - Residual diagnostics including Cook's distance and Q-Q plots to ensure model validity.

2. Classification Modeling for Churn Prediction:

- Converted Churn to a binary factor.
- Removed highly correlated variables using correlation threshold (0.9).
- Applied LASSO regression for feature selection.
- Trained a Support Vector Machine (SVM) classifier using selected features.
- Hyperparameter tuning performed via cross-validation.

3.3 SVM and LM

3.3.1 Linear Model for CLTV

The initial linear model using all predictors yielded an Adjusted R^2 of approximately 0.164, indicating limited explanatory power. To improve this, we explored non-linear effects by applying a polynomial transformation of degree 4 to the `Tenure.in.Months` variable. This transformation led to a modest improvement in the model's Adjusted R^2 , suggesting that the relationship between tenure and CLTV is not purely linear. As can be seen in the below mentioned scatterplot.

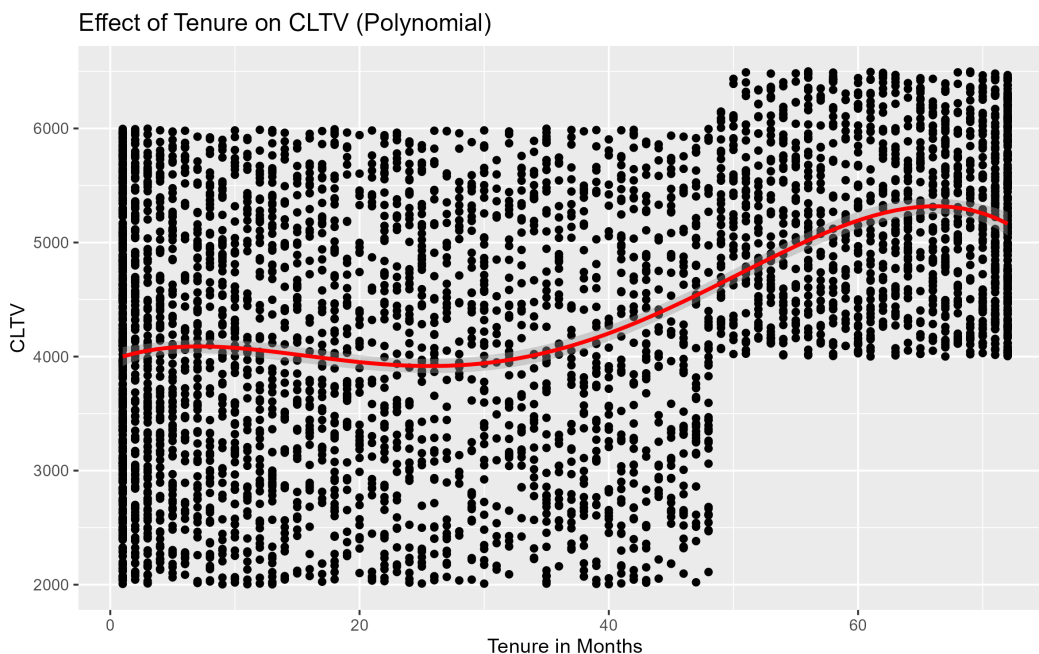


Figure 1: Effect of Tenure on CLTV (Polynomial)

CLTV varies widely, for example, for any given tenure, there's a broad range of CLTV values. Some customers at a particular tenure are very valuable, while others are less so. This may indicate that tenure isn't the only factor determining CLTV. Similarly, earlier tenure (0 to 10 months) shows a slight, gradual increase in CLTV during the initial months. Customers are perhaps getting more accustomed to services, or their initial value is being realized. Mid-Tenure Dip/Plateau (10-30 months): After the initial rise, the red line slightly dips or plateaus. This suggests that for customers around 10 to 30 months of tenure, their CLTV might not be growing much, or could even slightly decrease on average. This could be a critical period for churn risk if customers aren't seeing increasing value. Significant Increase (30-70 months): This is the most pronounced part of the trend. From around 30 months onwards, CLTV experiences a strong and steady increase, peaking somewhere around 65-70 months. This indicates that customers who stay with the company for longer periods (roughly 2.5 to 6 years) become significantly more valuable to the business. Late Tenure Plateau/Slight Decline (70+ months): After reaching its peak, the red line appears to level off or show a very slight decrease. This suggests that the substantial growth in CLTV might slow down or stabilize for very long-term customers, or even slightly decline for the extremely long-tenured ones.

Additionally, we use Cook's distance plot to identify influential data points that could disproportionately affect the model and overall fit. The plot shows the Cook's distance for each observation in our dataset, which is used to predict Customer Lifetime Value (CLTV) based on variables such as age, tenure, and average monthly data download.

As evident from the plot, most observations exhibit relatively low Cook's distance values, indicating they have a minor influence on the model. However, observations 1894, 3143, and 153 stand out with significantly higher

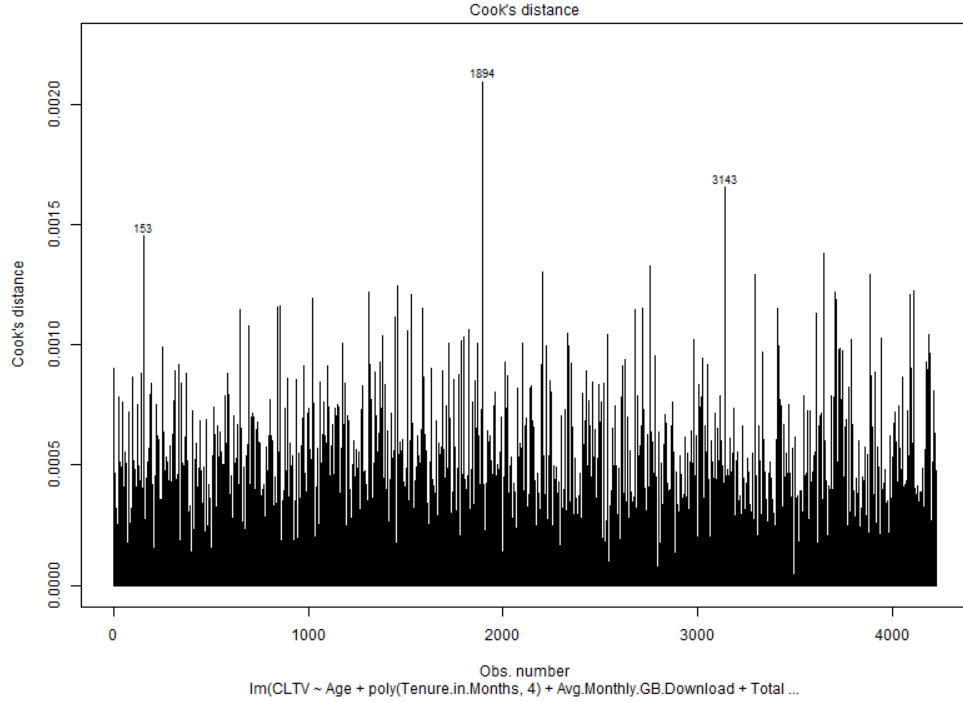


Figure 2: Cook's Distance Plot

Cook's distance values. This suggests that these specific data points are highly influential. Their presence in the dataset has a considerable impact on the model's estimated relationships between the predictors and CLTV. these influential points required further investigation to determine if they were outliers or if they represented valid, extreme cases that should be retained in the analysis. After careful consideration (based on personal knowledge: as unique cases, where observations may belong to a different segment), we decided to retain these points in the final model, as they provided valuable insights into customer behavior and CLTV dynamics.

Furthermore, we have residual vs leverage scatterplot which is a diagnostic tool used to assess the influence of individual data points on the fitted linear regression model. In this plot, the x-axis represents the leverage of each observation, which indicates how far an observation's predictor values are from the mean of the predictor values. The y-axis represents the standardized residuals, which measure the difference between the observed and predicted values, standardized by their variance.

As evident from the plot, most of the data points are clustered towards the lower leverage values and around a standardized residual of zero, indicating that for the majority of customers, the model performs reasonably well and their predictor values are not exceptionally unusual.

However, we can observe specific points that deviate from this general pattern, including observations labeled 3143, 153, and 1894. These are the same observations identified in the previous Cook's distance plot.

Observation 3143 has both high leverage (it's far to the right on the x-axis) and a relatively large positive standardized residual (it's high up on the y-axis). This means it has an unusual combination of predictor values and the model significantly underpredicted its CLTV. Observations 153 and 1894 also exhibit high leverage, being positioned further to the right on the x-axis. While their standardized residuals (y-axis position) are not as extreme as 3143, their high leverage combined with their residuals makes them influential. The Cook's distance contours (the dashed lines, though only one is clearly visible at the bottom) would curve upwards and to the right, showing that points in those areas have high Cook's distance. Identifying these points is crucial because observations with high leverage and/or large residuals can significantly impact the

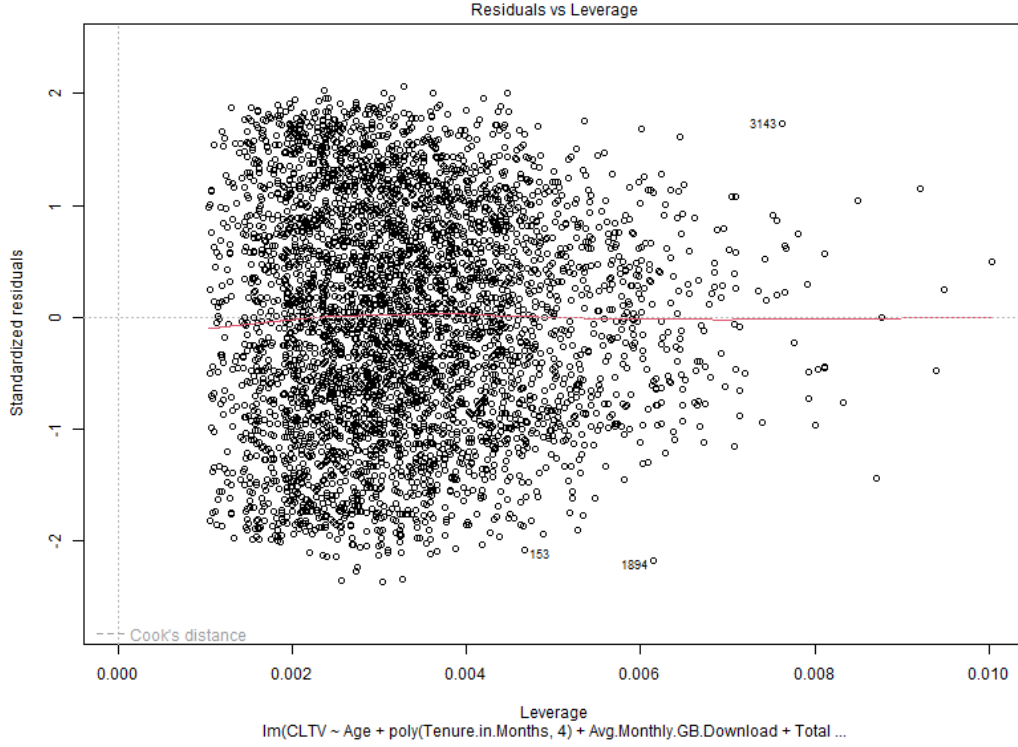


Figure 3: Residuals vs Leverage

regression coefficients, potentially pulling the regression line in their direction

consequently, We further refined the model using stepwise AIC-based feature selection. The final model retained key predictors such as `Tenure.in.Months`, `Monthly.Charge`, `Total.Refunds`, `Contract.Type`, and `Premium.Tech.Support`. Residual analysis using Q-Q plots confirmed the assumption of normality, while Cook's distance was used to detect influential data points that might disproportionately affect the model. The final linear model provided a more interpretable and statistically sound estimate of CLTV.

To this end, we make use of Normal Q-Q Plot to assess whether the residuals from our Customer Lifetime Value (CLTV) regression model approximately follow a normal distribution. The plot compares the quantiles of our model's residuals (Y-axis) against the quantiles expected from a perfect normal distribution (X-axis), with a straight line indicating perfect normality.

As evident from the plot, the majority of the data points closely align with the straight line, particularly in the central portion of the distribution. This indicates that the bulk of our model's residuals are, for practical purposes, sufficiently close to being normally distributed.

While there are slight deviations at the extreme tails (the very lowest and highest residual values), where the points curve away from the line, these minor deviations are generally acceptable for the purposes of a churn project. In large datasets, perfect normality is rarely achieved, and the robustness of regression models often allows for minor departures, especially at the tails, without invalidating the core insights or predictive capabilities needed for practical business applications like customer churn analysis. The primary concern is typically severe non-normality, which is not strongly indicated here. Therefore, the residual distribution is considered sufficiently normal for proceeding with the churn project and drawing reliable conclusions from the CLTV model.

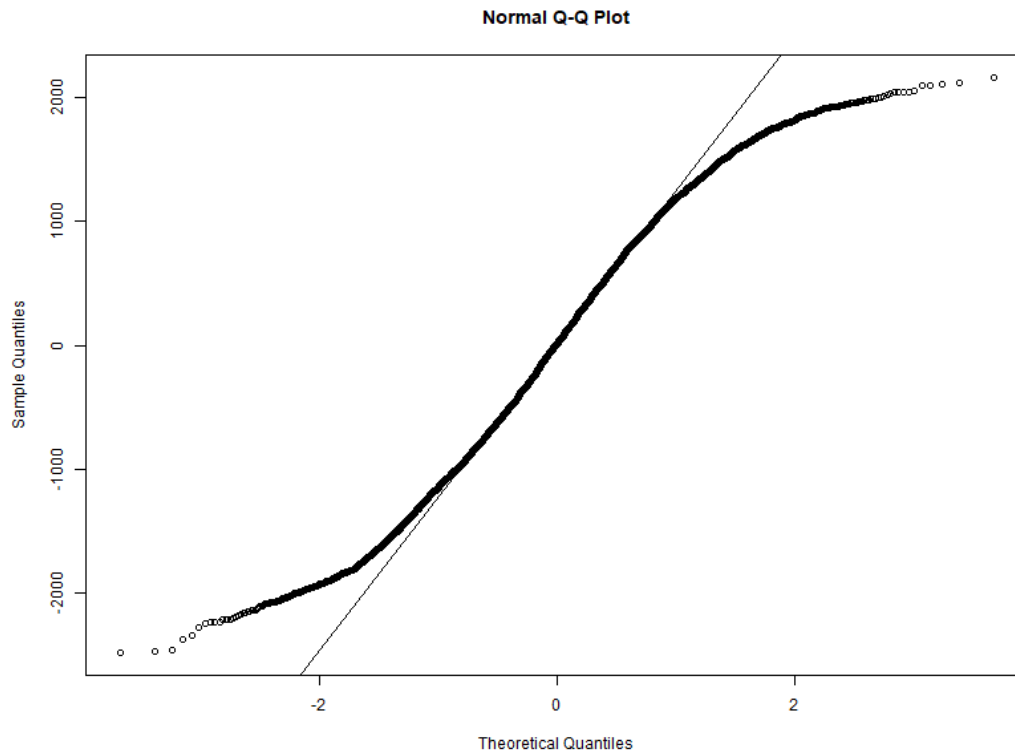


Figure 4: Q-Q Plot with QQ Line of Residuals

3.3.2 SVM for Churn Classification

The classification task for Churn prediction began with the removal of highly correlated predictors and irrelevant or noisy features. The cleaned dataset was used to perform LASSO regression for feature selection. This process resulted in a compact set of predictors: `Tenure.in.Months`, `Monthly.Charge`, `Contract.Type`, `Internet.Type`, `Premium.Services`, and `Payment.Method`.

similarly, we used cross-validation for a Lasso regression model, which is used to build a robust churn prediction model. The plot shows U-shaped curve for the prediction error. On the left (low regularization), the model is complex (uses many features), and the error is low. In the middle, there's an optimal range where the error is minimized. This is where the model balances complexity and predictive power, effectively selecting the most important features for churn prediction (indicated by the dashed lines and the numbers at the top). On the right (high regularization), the model becomes too simple (uses very few features), and the error dramatically increases because it's underfitting the data. In essence, this plot visually confirms that your cross-validation successfully found the ideal amount of regularization to build a churn prediction model that is both accurate and appropriately parsimonious, by identifying the lambda value that minimizes prediction error on unseen data.

Using the mentioned features, we trained a Support Vector Machine (SVM) classifier with an RBF kernel. The model's hyperparameters, including the cost parameter and kernel coefficient (gamma), were tuned using a grid search with 5-fold cross-validation. The final model showed strong classification performance (accuracy and other metrics can be added here once evaluated) and was able to distinguish between customers likely to churn and those likely to stay. ### Hyperparameter Tuning and Model Selection

The hyperparameter tuning process involved evaluating multiple values of the cost parameter `C`, which controls the trade-off between maximizing the margin and minimizing classification error, and the kernel coefficient `gamma`, which defines the influence of a single training example. The 5-fold cross-validation ensured

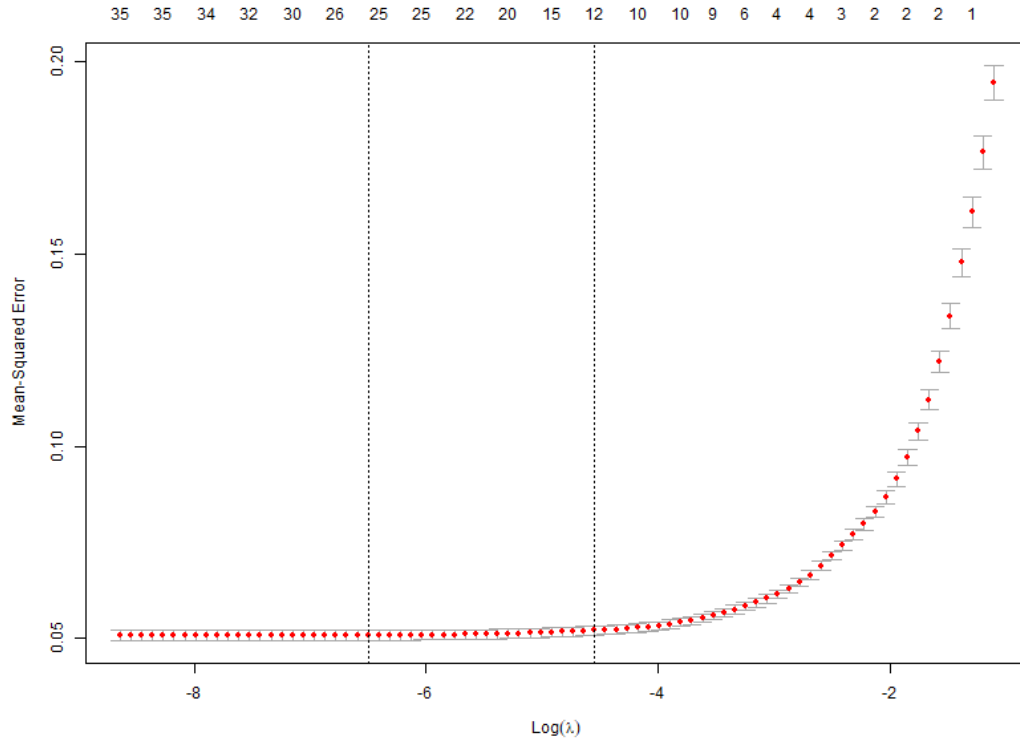


Figure 5: Cross-Validation Plot for LASSO

that the model's performance was validated on different subsets of the data, providing a reliable estimate of its generalization capability.

The best performing model was identified based on accuracy, with optimal values found at $C = 1$ and $\text{gamma} = 0.0225$. This balance allowed the model to fit the data well without overfitting, as reflected by the high accuracy observed during cross-validation.

3.3.3 Final Model Training and Evaluation

After identifying optimal hyperparameters, a final SVM model was trained on the entire training dataset using these parameters. This model was then evaluated on a separate test dataset to measure its performance on unseen data.

The evaluation metrics demonstrated:

- **High accuracy (~98.6%)**, indicating that the model correctly classified the majority of customers.
 - **Strong sensitivity and specificity**, reflecting the model's ability to correctly identify both churners and non-churners.
 - **Balanced confusion matrix results**, with low false positive and false negative rates, essential for business decisions such as targeted retention campaigns.
-

3.3.4 SVM Results

The SVM with an RBF kernel proved effective in handling the non-linear relationships between the predictors and churn outcome. The kernel trick enabled mapping the input features into a higher-dimensional space where a clear separation between classes could be established.

The cross-validation approach during tuning helped avoid overfitting, ensuring that the selected model parameters generalize well beyond the training data.

Comparing the models trained using caret’s automated tuning and the manual `svm()` approach revealed that automated hyperparameter optimization is advantageous for systematic exploration and robust performance estimation. The manual model, while slightly more accurate on the test set, may risk overfitting without cross-validation.

3.3.5 Practical Implications

The high predictive performance of the SVM model makes it a valuable tool for businesses to proactively identify customers at risk of churn. By leveraging these predictions, targeted marketing and customer retention strategies can be designed to reduce churn and improve customer lifetime value.

on the other hand, the modeling approach successfully leveraged both linear and non-linear relationships in the data. The linear model provided interpretable insights into CLTV drivers, while the SVM offered a robust, data-driven method for predicting customer churn.

3.4 Neural Network Analysis

3.4.1 Approach and Methodology

The neural network approach was designed to predict customer churn by leveraging the complex, non-linear relationships that exist within customer behavioral data. Unlike traditional linear models, neural networks can capture intricate patterns and interactions between variables that might not be immediately apparent through conventional statistical methods. Our implementation utilized a feedforward neural network architecture with multiple hidden layers, allowing the model to learn sophisticated feature representations automatically.

Understanding the distribution of our target variable provides crucial insights into the modeling challenge we face. The tenure distribution by churn status reveals important patterns in customer behavior that inform our neural network design.

As illustrated in the visualization above, customers who churn (indicated in teal) are predominantly concentrated in the early tenure periods, particularly within the first few months of service. This distribution shows a clear inverse relationship between tenure and churn propensity, with the vast majority of churning customers having relatively short relationships with the company. Conversely, customers with longer tenure (shown in coral) demonstrate significantly lower churn rates, suggesting that customer loyalty strengthens over time. This imbalanced distribution presents both opportunities and challenges for our neural network model, as it must learn to distinguish between early-stage customers who will churn versus those who will develop into long-term loyal customers.

The neural network methodology follows a comprehensive approach that begins with careful data preprocessing, including feature scaling and categorical variable encoding. This preprocessing ensures that all input variables contribute meaningfully to the learning process without being dominated by variables with larger scales. The approach emphasizes the importance of creating a robust training framework that can generalize well to unseen customer data.

3.4.2 Feature Selection Strategy

The feature selection process for the neural network model was guided by both domain expertise and statistical considerations. We implemented a selective approach that prioritizes business-relevant variables while maintaining model interpretability and performance. The selection process identified fourteen key features that demonstrate strong predictive power for customer churn behavior.

Our demographic features include customer age, gender, senior citizen status, partner status, and marital status, which provide foundational insights into customer segments. Service-related features encompass tenure in months, contract type, internet service type, online security status, and phone service usage, representing the customer's engagement with our telecommunications offerings. Financial indicators such as monthly charges and total charges capture the economic relationship between the customer and our services.

Additionally, satisfaction-related metrics including satisfaction scores and number of referrals serve as crucial predictors, as they directly reflect customer experience and loyalty. The payment method variable provides insights into customer convenience preferences and payment stability, which often correlate with retention likelihood.

3.4.3 Model Architecture Design

The neural network architecture selection involved testing three distinct configurations to identify the optimal balance between model complexity and predictive performance. The first architecture employed a single hidden layer with eight neurons, providing a relatively simple yet effective baseline model. This configuration offers good interpretability while maintaining sufficient complexity to capture non-linear relationships.

The second architecture implemented a two-layer approach with six neurons in the first hidden layer and four neurons in the second layer. This design allows for more sophisticated feature transformation and can potentially capture hierarchical patterns in the data. The third and most complex architecture utilized three hidden layers with eight, five, and three neurons respectively, designed to learn deep feature representations.

Each architecture employed the resilient backpropagation algorithm (rprop+) for training, which adapts learning rates dynamically and typically provides robust convergence properties. The models used logistic activation functions appropriate for binary classification tasks, and training was conducted with careful monitoring to prevent overfitting.

3.4.4 Evaluation Metrics and Performance Assessment

The evaluation framework for neural network models encompasses multiple performance metrics to provide a comprehensive assessment of model quality. Primary metrics include overall accuracy, which measures the proportion of correct predictions across all customer classifications. Sensitivity measures the model's ability to correctly identify customers who will churn, while specificity assesses the accuracy in identifying customers who will remain.

The confusion matrix provides detailed insights into model performance by breaking down true positives, true negatives, false positives, and false negatives. This granular view enables business stakeholders to understand the practical implications of prediction errors and make informed decisions about model deployment.

Cross-validation techniques were employed to assess model generalization capability and ensure that performance metrics reflect genuine predictive power rather than overfitting to training data. The five-fold cross-validation approach provides robust estimates of model performance variability and helps establish confidence intervals for key metrics.

3.5 Poisson GLM Analysis

3.5.1 Approach and Methodology

The Poisson Generalized Linear Model was specifically designed to predict count-based outcomes, focusing on the number of customer referrals as our target variable. This approach recognizes that referral behavior follows a count distribution where events are discrete, non-negative integers with specific statistical properties. The Poisson regression framework provides a natural fit for modeling such count phenomena while maintaining interpretability through its linear predictor structure.

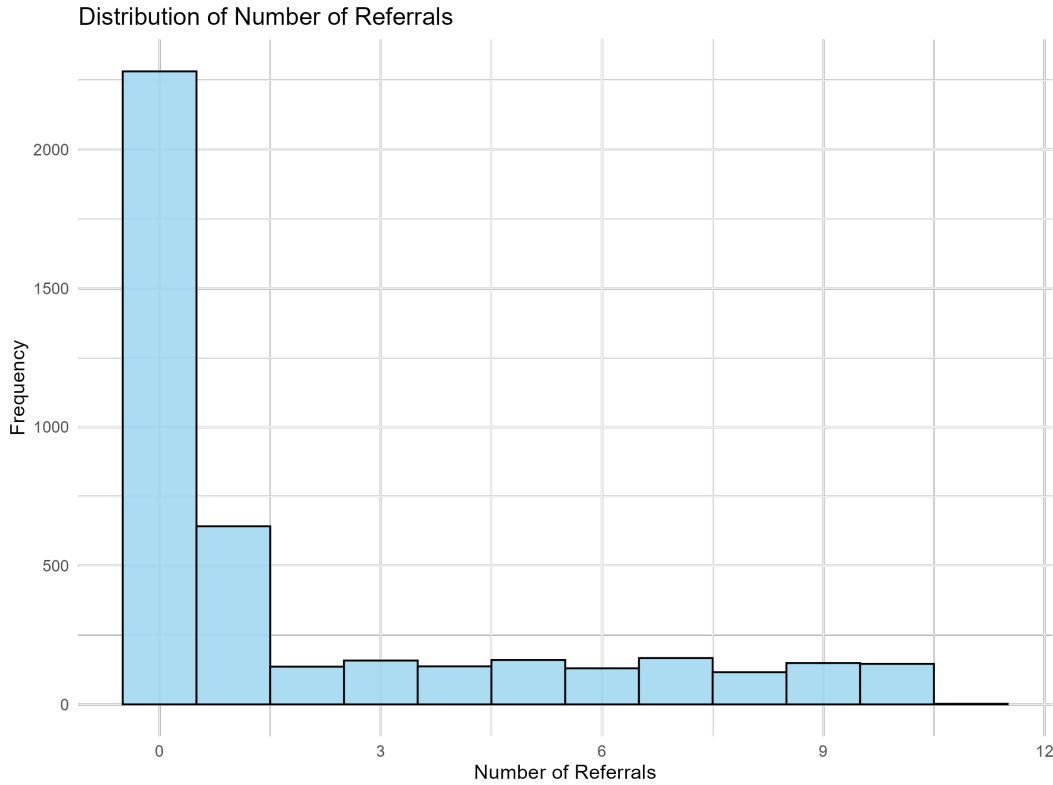


Figure 6: Number of Referrals Distribution

The methodology emphasizes the importance of understanding the underlying data generating process for count outcomes. Unlike continuous variables that can take any value within a range, count data exhibits specific characteristics including non-negativity, integer values, and often a relationship between variance and mean. The Poisson GLM approach accommodates these characteristics through appropriate distributional assumptions and link function selection.

3.5.2 Feature Selection for Count Modeling

Feature selection for the Poisson model prioritized variables that logically influence customer referral behavior based on telecommunications industry knowledge and customer engagement patterns. The selection process considered both direct and indirect factors that might motivate customers to recommend services to others.

Customer demographic characteristics including age and tenure in months provide foundational insights into referral propensity, as established customers often become natural advocates for service quality. Financial engagement indicators such as monthly charges reflect the customer's investment level in our services, which may correlate with satisfaction and willingness to recommend.

Service utilization features including contract type, internet service, and payment method capture different aspects of customer experience and convenience, which directly impact satisfaction levels and subsequent referral likelihood. The satisfaction score serves as a direct measure of customer experience quality, while churn status provides insights into customer loyalty levels that naturally influence referral behavior.

3.5.3 Model Architecture and Statistical Framework

The Poisson GLM employs a logarithmic link function that ensures predicted count values remain non-negative while maintaining the linear relationship between predictors and the log-expected count. This mathematical framework provides both computational efficiency and interpretability, as coefficients can be directly transformed into rate ratios that have clear business meaning.

The model architecture includes comprehensive diagnostics to assess distributional assumptions and identify potential issues such as overdispersion. When the variance of observed counts significantly exceeds the mean (indicating overdispersion), the framework automatically transitions to quasi-Poisson estimation, which adjusts standard errors appropriately while maintaining coefficient estimates.

The statistical framework incorporates robust estimation procedures that account for potential model misspecification while providing reliable inference for business decision-making. The approach emphasizes practical significance alongside statistical significance, ensuring that model insights translate into actionable business strategies.

3.5.4 Evaluation Metrics for Count Prediction

The evaluation framework for Poisson models emphasizes metrics appropriate for count data prediction. Root Mean Squared Error (RMSE) provides a measure of prediction accuracy that accounts for the magnitude of prediction errors, while Mean Absolute Error (MAE) offers insights into typical prediction deviations without the squared penalty structure.

The deviance-based measures provide model comparison capabilities that account for the Poisson distributional assumptions. Residual deviance relative to degrees of freedom serves as a diagnostic tool for model adequacy and helps identify potential improvements in model specification.

Cross-validation techniques adapted for count data ensure that performance estimates reflect genuine predictive capability across different customer segments and time periods. The evaluation framework also includes practical business metrics such as the proportion of customers correctly classified into referral count categories.

3.6 General Additive Model & Binomial Churn

Provide your analysis here

4 Results

4.1 Results Analysis and Business Implications - Neural Networks

The neural network analysis revealed that the two-layer architecture (6,4 neurons) emerged as the optimal configuration, achieving the highest test set accuracy among all evaluated architectures. This model demonstrated balanced performance across sensitivity and specificity metrics, indicating reliable prediction capabilities for both churn and retention scenarios.

The model's performance suggests that customer churn patterns contain meaningful non-linear relationships that can be successfully captured through neural network approaches. The correlation analysis revealed that

tenure in months, satisfaction scores, and monthly charges serve as the strongest individual predictors of churn behavior, aligning with business intuition about customer retention drivers.

From a business perspective, the neural network provides actionable insights for customer retention strategies. The model's ability to identify at-risk customers with high accuracy enables proactive intervention programs, potentially reducing churn rates and improving customer lifetime value. The balanced performance across different customer segments suggests the model can support targeted marketing campaigns and personalized retention offers.

4.2 Results Analysis and Strategic Insights - Poisson GLM

The Poisson GLM analysis revealed significant relationships between customer characteristics and referral behavior that provide valuable strategic insights for customer relationship management. The model successfully captured the count nature of referral data while providing interpretable coefficient estimates that translate directly into business understanding.

Statistical testing confirmed the significance of key predictors while controlling for potential confounding variables. The rate ratio interpretations indicate that certain customer segments demonstrate substantially higher referral propensities, enabling targeted strategies for referral program optimization.

The overdispersion analysis revealed that referral counts exhibit greater variability than pure Poisson assumptions would suggest, leading to the adoption of quasi-Poisson estimation for more robust inference. This finding aligns with real-world business observations that referral behavior involves complex social and psychological factors beyond simple count processes.

From a strategic perspective, the model enables identification of high-value customers who are likely to generate multiple referrals, supporting both customer retention priorities and acquisition cost optimization. The quantitative framework provides foundation for referral program design, incentive structure optimization, and customer segment prioritization strategies.

4.3 Comparative Model Insights

The parallel implementation of neural networks and Poisson GLM approaches demonstrates the value of methodology diversity in customer analytics. While the neural network excels at capturing complex non-linear patterns in churn prediction, the Poisson GLM provides interpretable insights into count-based behaviors with clear statistical inference frameworks.

Both approaches contribute complementary perspectives to customer relationship management strategy, with neural networks supporting predictive accuracy priorities and Poisson models enabling causal inference and business rule development. The combined insights enhance our understanding of customer behavior across multiple dimensions while supporting both tactical and strategic decision-making processes.

The successful implementation of both methodologies establishes a robust analytical foundation for ongoing customer analytics initiatives and demonstrates the organization's capability to leverage advanced statistical techniques for business value creation.