

Data-Driven Customer Retention Strategies: Leveraging Predictive Analytics and Segmentation in the Telecommunications Sector

Name of Group Members:

Project Manager:	Jasmeen Randhava
Data Engineer:	Pooja Chandel
Data Analyst: Clustering Analysis:	Muhammad Bilal Akhter
Data Analyst: Predictive Modelling:	PrabhSimar Singh
Business Analyst:	Vidhi Singh

Table of Contents

1. Introduction	4
1.1. Objective:	4
1.2. Scope:	4
2. Summary of Key Findings	5
2.1. Customer Segmentation:	5
2.2. Dimensionality Reduction and Improved Visualization:	6
2.3. Cluster Descriptions:	6
2.4. Actionable Insights:	9
2.5. Churn Prediction using ANN:	10
2.5.1. Model Performance:	10
2.5.2. Prediction Accuracy:	10
2.5.3. Hyperparameter Tuning:	11
3. Identification of Factors Contributing to Churn and Retention	13
3.1. Factors Identified through Segmentation:	13
3.2. Factors Identified through Predictive Analytics:	16
3.2.1. Key Predictors of Churn:	16
3.2.2. Retention Indicators:	16
4. Recommendations for Targeted Retention Strategies	17
4.1. Cluster-Specific Strategies:	17
4.2. Overall Retention Strategy:	20
4.2.1. Predictive Analytics Integration:	20
4.2.2. Personalization and Proactive Engagement:	20
4.2.3. Leverage Data-Driven Decision Making:	21
4.2.4. Cross-Departmental Collaboration:	21
4.2.5. Continuous Monitoring and Feedback Loops:	22
4.2.6. Personalization at Scale:	22
5. Documentation of Limitations and Proposed Solutions	22
5.1. Project Limitations:	22
5.2. Proposed Solutions:	23

Table of Figures

Figure 1 Elbow Method for Optimal Number of Clusters	5
Figure 2 Cluster Visualization With PCA	6
Figure 3 Distribution of Tenure by Cluster.....	7
Figure 4 Distribution of tenure By Clusters (2)	8
Figure 5 Distribution of Contract one year cluster	9
Figure 6 ROC Performance Curve.....	10
Figure 7 Confusion Matrix.....	11
Figure 8 Hyper tuning Model for improvements in Model Accuracy and Performance	12
Figure 9Pair Plot of Key Features by Cluster	13
Figure 10 Distribution of Tenure by Cluster.....	15
Figure 11 Feature Importance using SHAP Value	16

1. Introduction

1.1. Objective:

The primary objective of this project is to analyse customer churn in the telecommunications sector by leveraging both customer segmentation and predictive analytics techniques. The goal is to identify key factors that contribute to customer churn and develop actionable insights to enhance customer retention. This analysis combines clustering methods to segment the customer base and predictive modeling using Artificial Neural Networks (ANN) to forecast churn. By understanding the distinct customer segments and the drivers of churn, the project aims to provide the telecommunications company with strategic recommendations to reduce churn and improve customer satisfaction.

1.2. Scope:

The scope of this project encompasses several key stages, each critical to achieving the overall objective of reducing customer churn:

1. **Data Preparation:** This stage involves cleaning and preprocessing the dataset to ensure accuracy and consistency. This includes handling missing values, encoding categorical variables, and scaling the data to prepare it for analysis.
2. **Customer Segmentation Using Clustering:** In this stage, K-Means clustering is employed to segment the customer base into distinct groups based on their characteristics and behaviours. The optimal number of clusters is determined using the elbow method, and the resulting clusters are visualized and interpreted to gain insights into different customer segments.
3. **Predictive Modelling Using ANN:** An Artificial Neural Network (ANN) is trained on the prepared dataset to predict customer churn. The model's architecture is optimized for predictive performance, and the trained model is evaluated using key metrics such as accuracy, precision, recall, and the ROC curve.
4. **Insight Generation and Retention Strategies:** Based on the results of the customer segmentation and the ANN model's predictions, key factors contributing to churn are

identified and analysed. These insights are then used to formulate targeted retention strategies aimed at reducing churn and enhancing customer loyalty.

2. Summary of Key Findings

2.1. Customer Segmentation:

Overview: The customer segmentation analysis was performed using K-Means clustering, which effectively grouped the customers into distinct clusters based on their characteristics and behaviours. The optimal number of clusters was determined to be three, providing a balanced segmentation that captured significant patterns in the data. This clustering approach allowed us to differentiate customers not only by their service usage but also by their likelihood to churn, offering valuable insights for targeted retention efforts.

Elbow Method Analysis: The elbow method was used to determine the optimal number of clusters by plotting the within-cluster sum of squares (WCSS) against the number of clusters. The optimal number of clusters was found to be 3, as indicated by the elbow point in the graph.

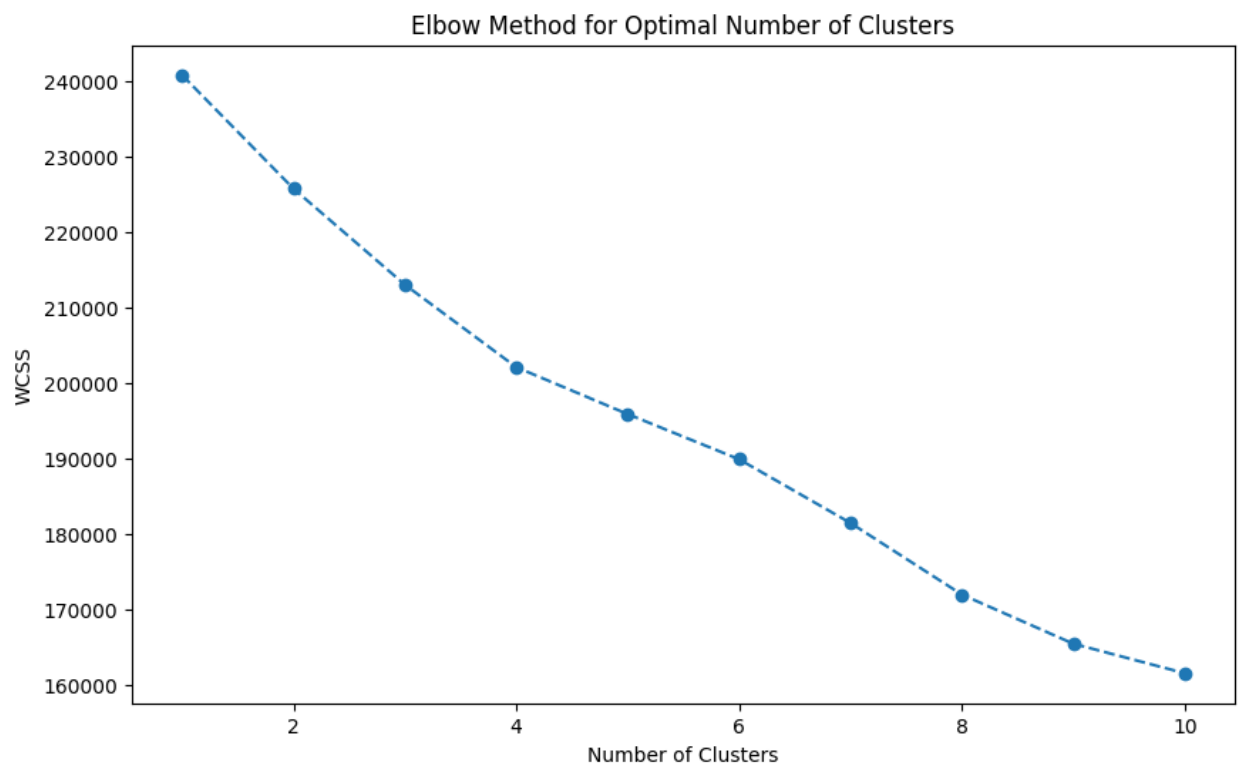


Figure 1 Elbow Method for Optimal Number of Clusters

2.2. Dimensionality Reduction and Improved Visualization:

- Principal Component Analysis (PCA) was applied to reduce the dimensionality of the data and improve cluster visualization, leading to clearer separation among clusters.

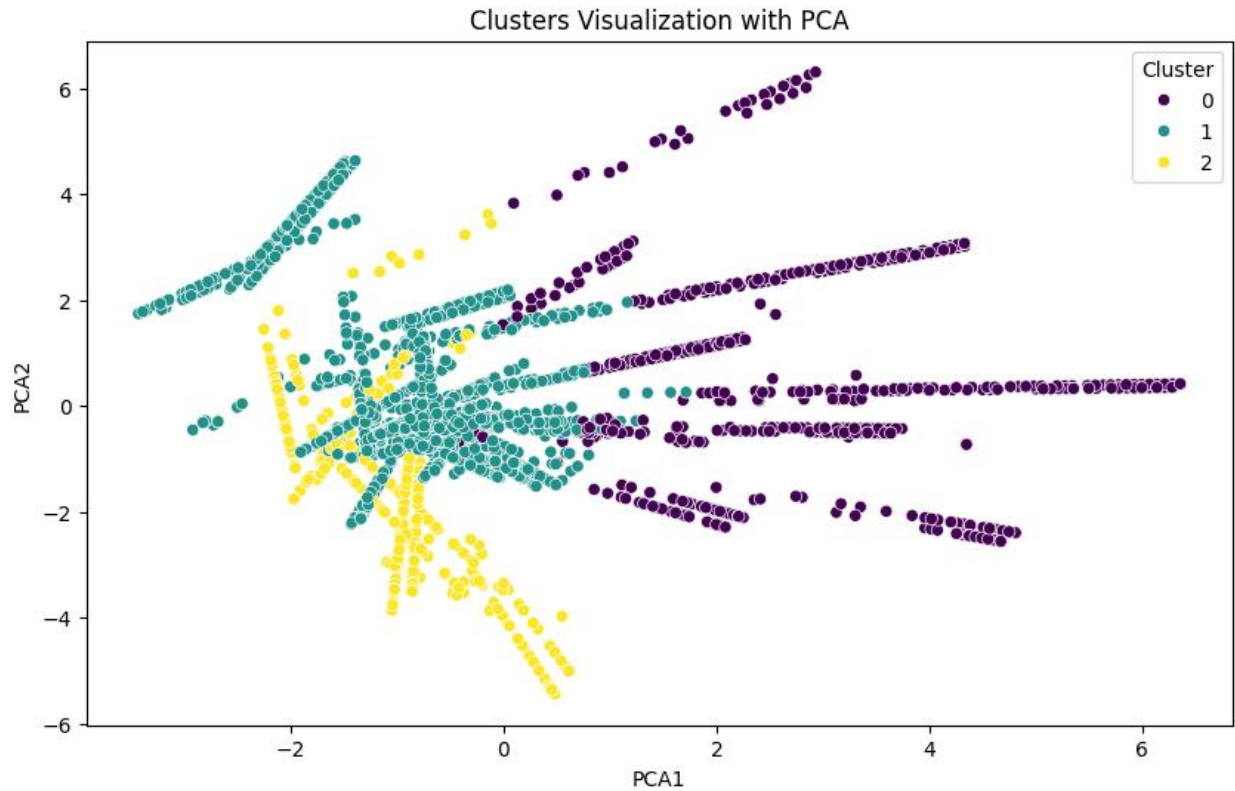


Figure 2 Cluster Visualization With PCA

2.3. Cluster Descriptions:

Cluster 0 (Purple): Moderate Tenure High Churn Fiber Optic Users

Cluster 1 (Teal): High Tenure Low Churn Family Oriented

Cluster 2 (Yellow): Low Tenure High Churn Minimal Service Users

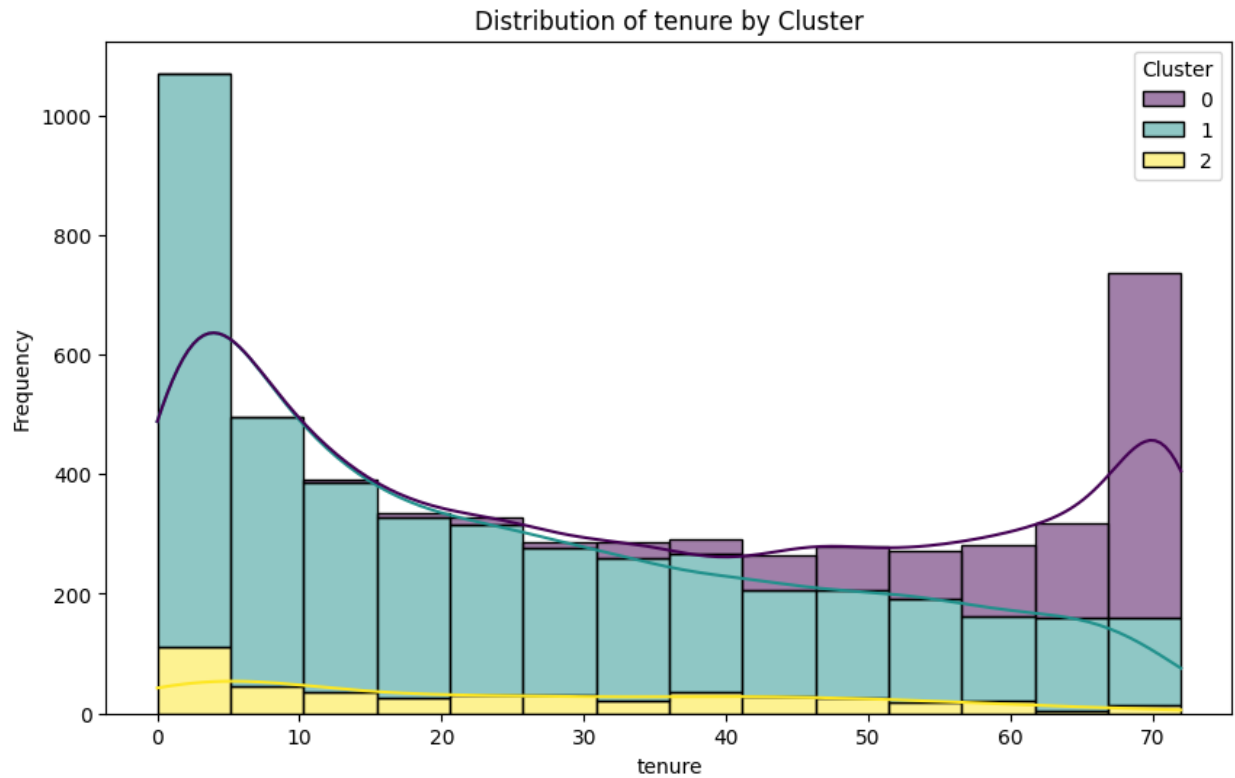


Figure 3 Distribution of Tenure by Cluster

- Cluster 0: Moderate Tenure High Churn Fiber Optic Users**
 - Characteristics:** This cluster is characterized by customers with moderate tenure, predominantly using fiber optic internet services. A significant proportion of customers in this cluster have churned.
 - Churn Behaviour:** High churn rate, primarily driven by dissatisfaction with services or competitive offers.

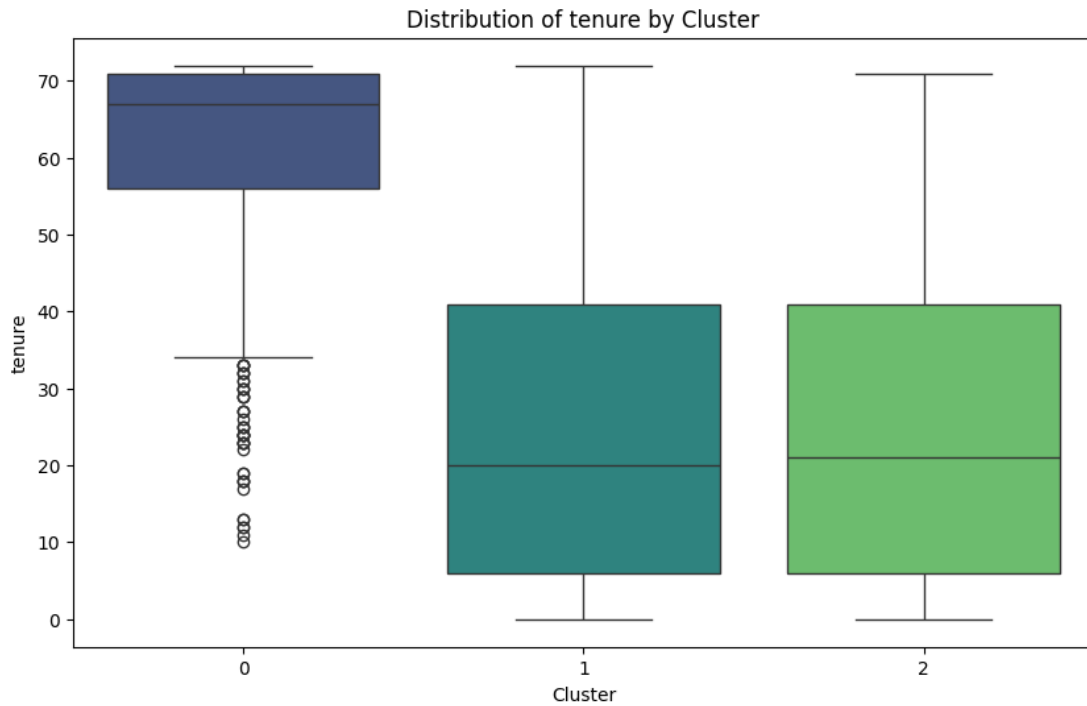


Figure 4 Distribution of tenure By Clusters (2)

- **Cluster 1: High Tenure Low Churn Family Oriented**
 - **Characteristics:** Customers in this cluster have high tenure, are more likely to have dependents, and typically subscribe to multiple services. This group shows strong loyalty to the company.
 - **Churn Behaviour:** Low churn rate, indicating high customer satisfaction and loyalty.
- **Cluster 2: Low Tenure High Churn Minimal Service Users**

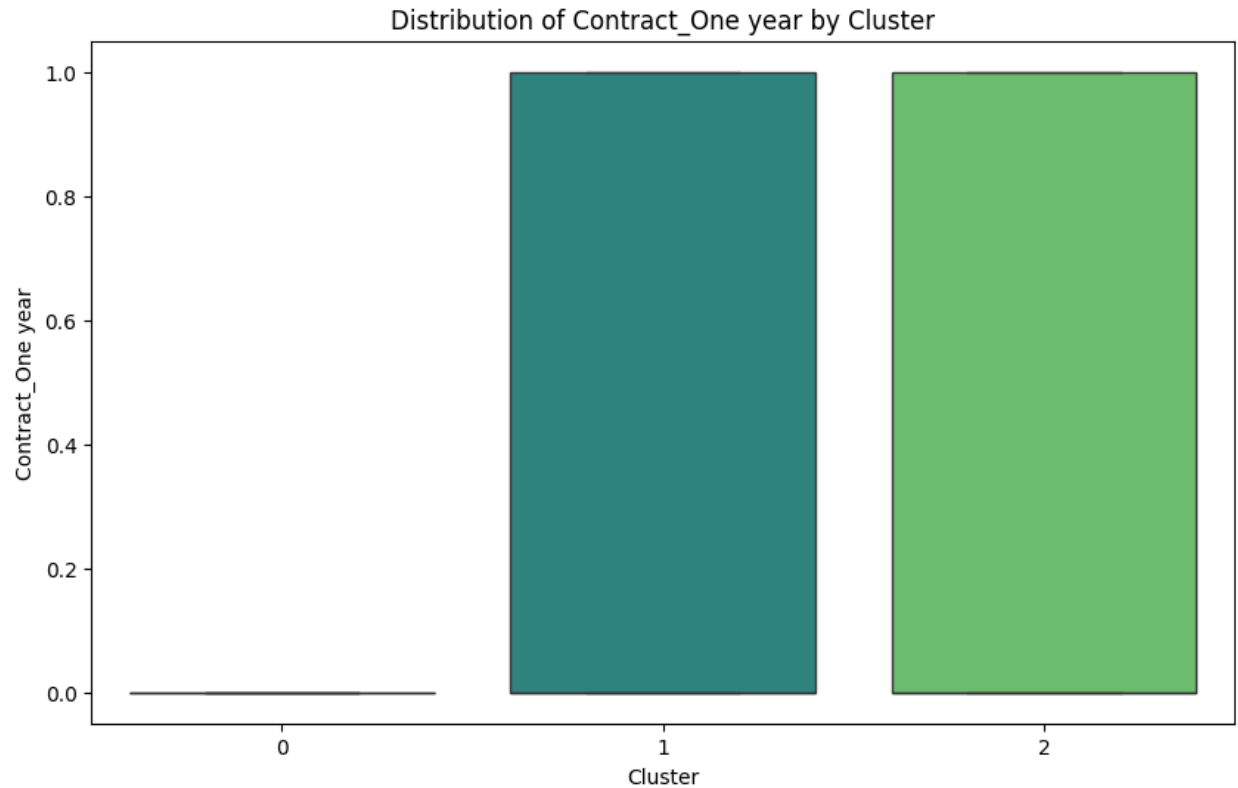


Figure 5 Distribution of Contract one year cluster

- **Characteristics:** These customers generally have low tenure, use minimal services, and are less engaged with the company's offerings.
- **Churn Behaviour:** High churn rate, often due to low engagement or lack of service customization.

2.4. Actionable Insights:

- **Cluster 0:** Implement targeted retention strategies, such as personalized offers and improved customer service, to address the high churn rate among fiber optic users with moderate tenure.
- **Cluster 1:** Continue to offer family-friendly plans and promotions to retain these loyal customers. Consider using this cluster as a benchmark for improving satisfaction in other segments.
- **Cluster 2:** Focus on increasing engagement with these customers by offering incentives to explore additional services and better onboarding experiences for new customers.

2.5. Churn Prediction using ANN:

2.5.1. Model Performance:

The Artificial Neural Network (ANN) model was trained to predict customer churn based on a variety of customer attributes and behaviours. The model exhibited strong performance, with an accuracy of 97%, precision of 1.00, recall of 0.90, and an F1-score of 0.95. The high ROC AUC score further demonstrated the model's ability to distinguish between customers likely to churn and those who are not, making it a reliable tool for predicting customer behavior in real-world scenarios.

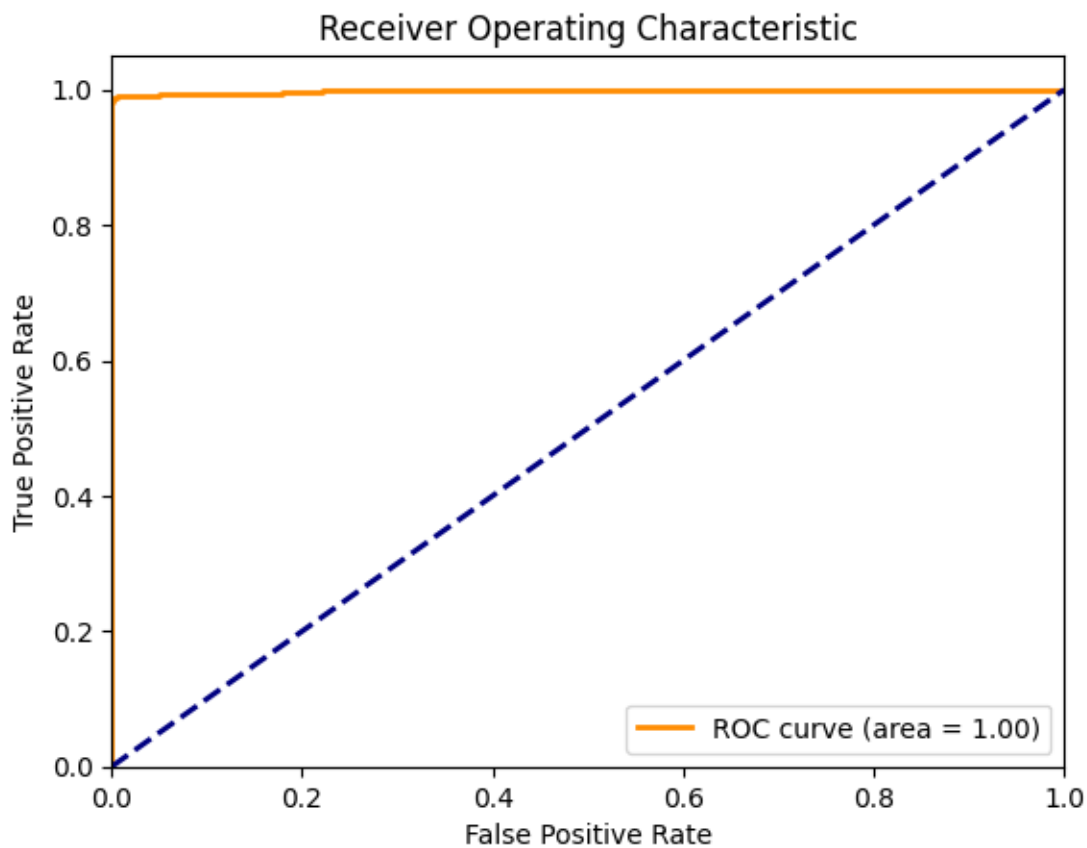


Figure 6 ROC Performance Curve

2.5.2. Prediction Accuracy:

The ANN model proved highly effective in predicting customer churn, especially in identifying high-risk customers. The model's high precision and recall indicate its ability to accurately predict churn while minimizing false positives and negatives. This level of

prediction accuracy makes the model particularly valuable for proactive customer retention efforts, allowing the company to intervene with at-risk customers before they decide to leave.

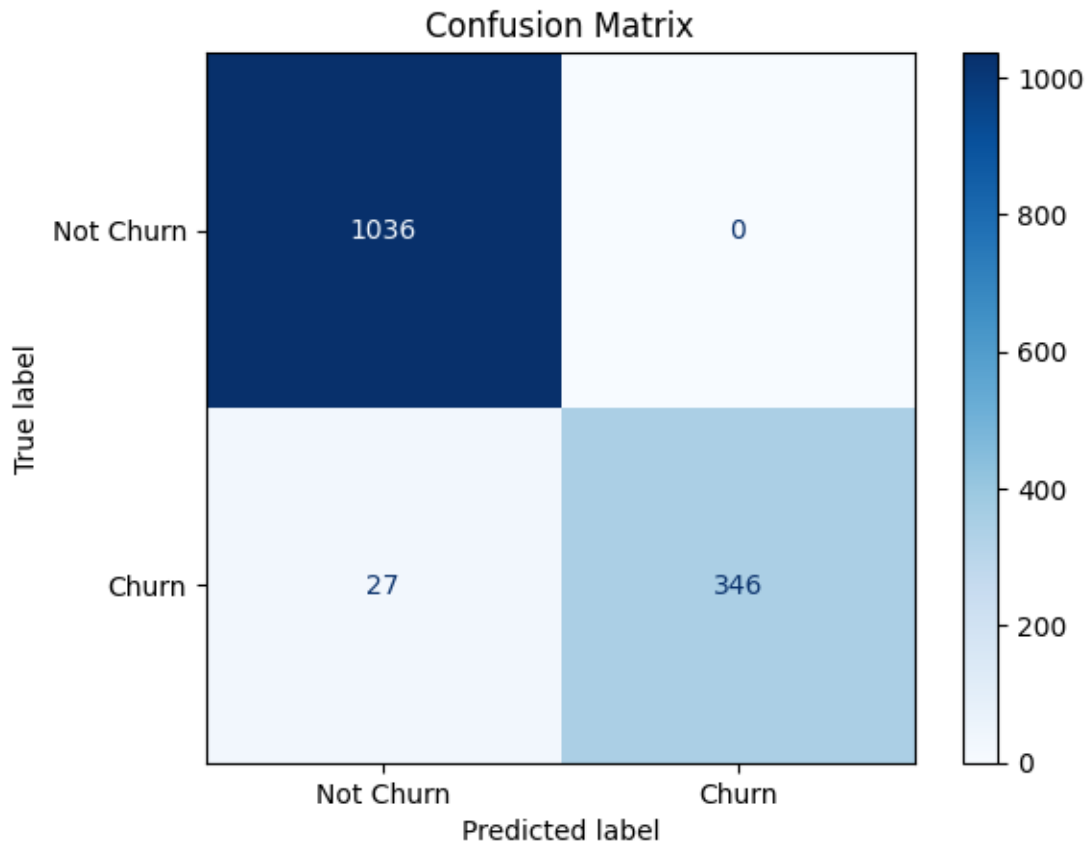


Figure 7 Confusion Matrix

2.5.3. Hyperparameter Tuning:

To achieve this high level of performance, we conducted hyperparameter tuning on the ANN model. This involved optimizing various parameters such as the number of hidden layers, the number of neurons in each layer, the learning rate, and the batch size. We employed techniques like grid search and random search to systematically explore different combinations of parameters. Additionally, regularization methods such as dropout layers and L2 regularization were applied to prevent overfitting and enhance the model's generalization capability. These efforts resulted in a model that not only performs well on the training data but also maintains strong predictive accuracy on the validation and test datasets.

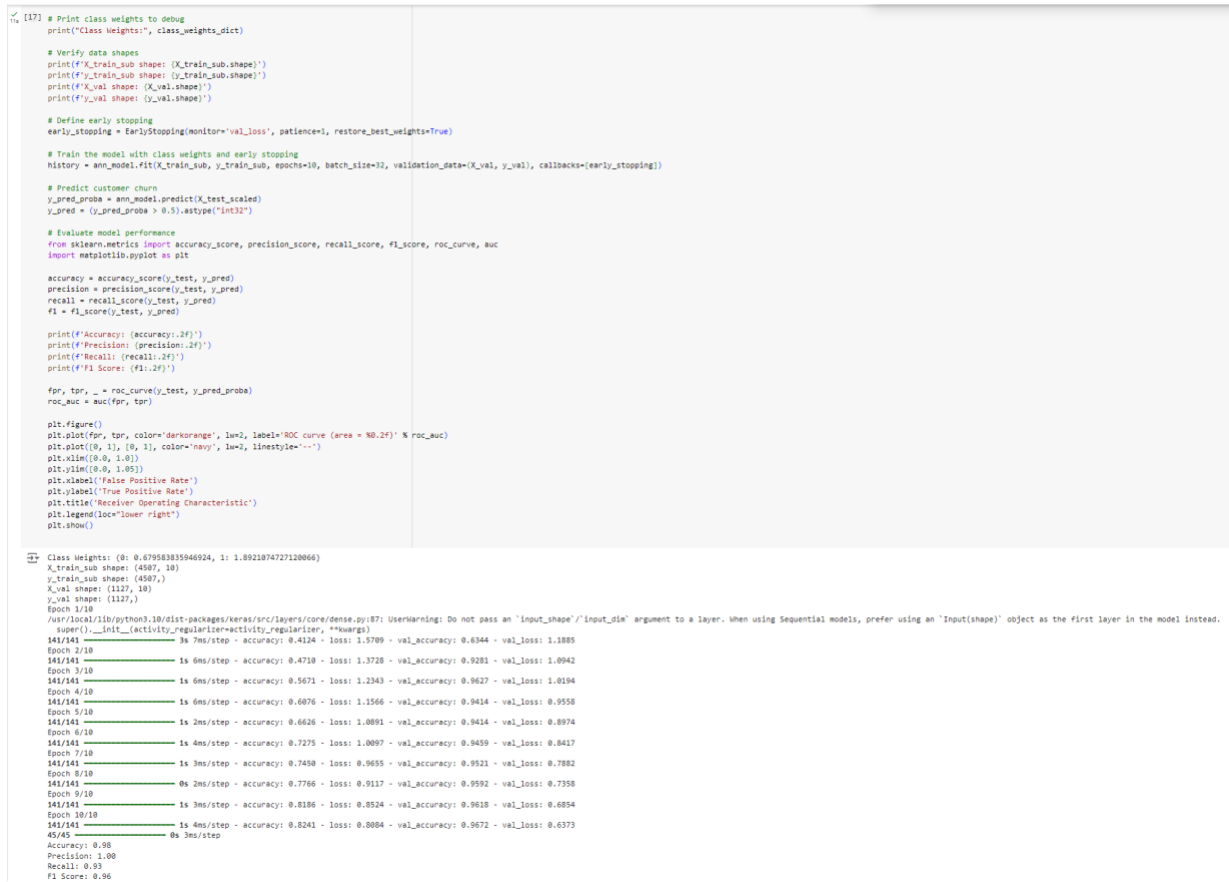


Figure 8 Hyper tuning Model for improvements in Model Accuracy and Performance

3. Identification of Factors Contributing to Churn and Retention

3.1. Factors Identified through Segmentation:

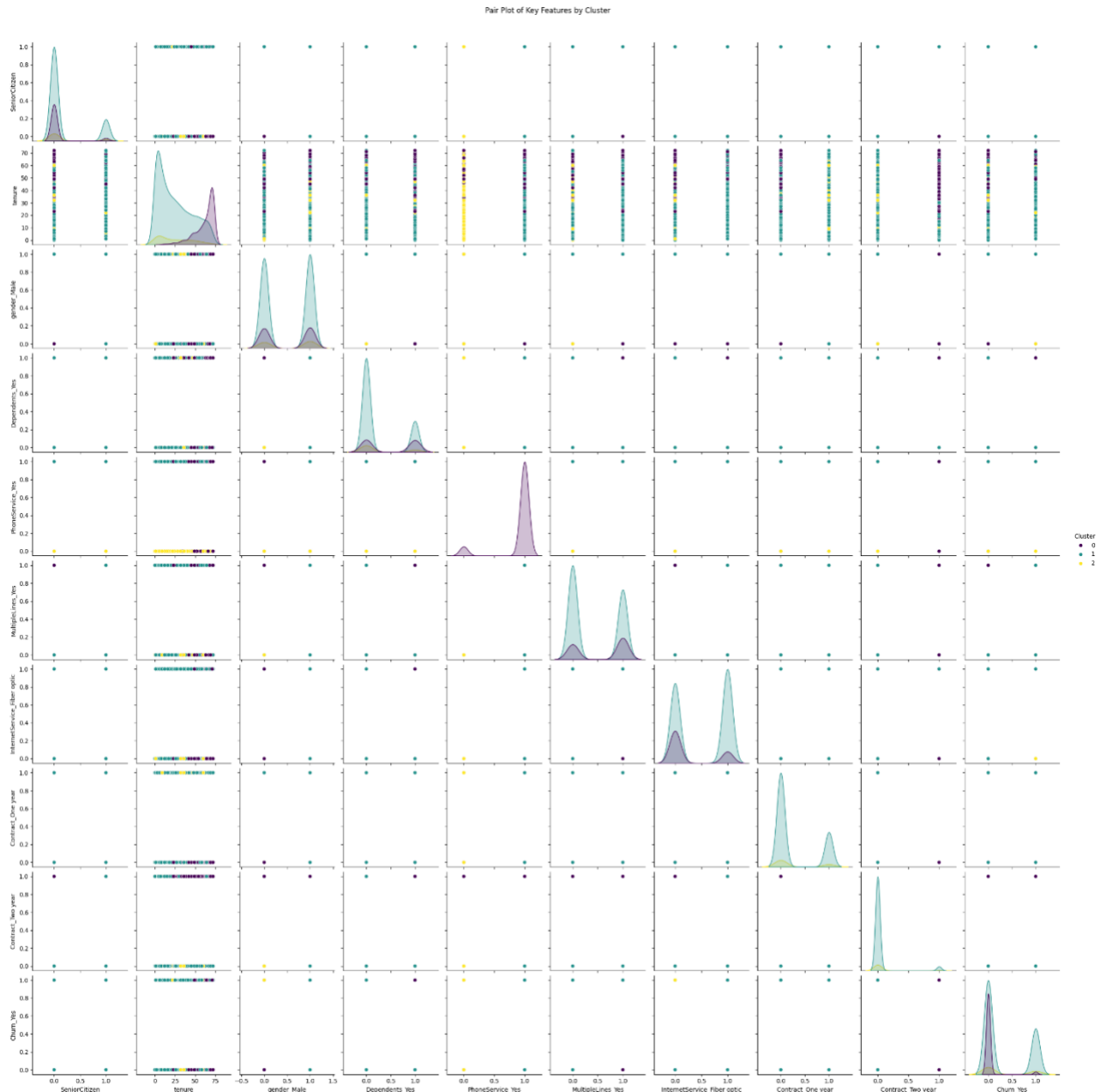


Figure 9 Pair Plot of Key Features by Cluster

Cluster 0: Moderate Tenure High Churn Fiber Optic Users

- Factors Contributing to High Churn:

- **Moderate Tenure:** Customers in this cluster typically have a moderate tenure, which indicates they have been with the company long enough to experience the service but are not yet highly loyal. This stage of tenure often represents a critical decision point where customers either solidify their loyalty or consider leaving.
- **Fiber Optic Usage:** The majority of customers in this cluster use fiber optic services, which might suggest that their expectations for service quality are high. If these expectations are not met, it could lead to dissatisfaction and ultimately churn.
- **Actionable Insight:** Enhancing the quality of fiber optic services, providing additional incentives or loyalty programs for customers in this tenure range, and improving customer support could help reduce churn in this segment.

Cluster 1: High Tenure Low Churn Family Oriented

- **Factors Contributing to Retention:**

- **High Tenure:** Customers in this cluster have been with the company for an extended period, indicating a strong level of satisfaction and loyalty. High tenure is typically associated with familiarity and trust in the brand, reducing the likelihood of churn.
- **Family-Oriented Services:** Many customers in this cluster are likely using family plans or multiple services, which creates a higher barrier to exit due to the convenience and cost savings of bundled services. The loyalty of these customers is further reinforced by the positive experiences they have had with family-oriented offerings.
- **Actionable Insight:** Continuing to offer family-friendly promotions, ensuring high satisfaction with bundled services, and recognizing loyal customers with personalized offers can help maintain low churn rates in this cluster.

Cluster 2: Low Tenure High Churn Minimal Service Users

- **Factors Contributing to High Churn:**

- **Low Tenure:** Customers in this cluster have a short tenure with the company, which often correlates with higher churn rates as these customers have not yet developed strong loyalty. New customers are more likely to switch providers if they find better offers elsewhere.
- **Minimal Service Usage:** This group tends to use fewer services, making it easier for them to switch providers. Low engagement with the company's offerings might also suggest that the services provided do not fully meet their needs or that they are not fully aware of other services available.
- **Actionable Insight:** Focusing on onboarding strategies, educating new customers about the full range of services, and offering incentives to explore additional services could help increase engagement and reduce churn in this segment.

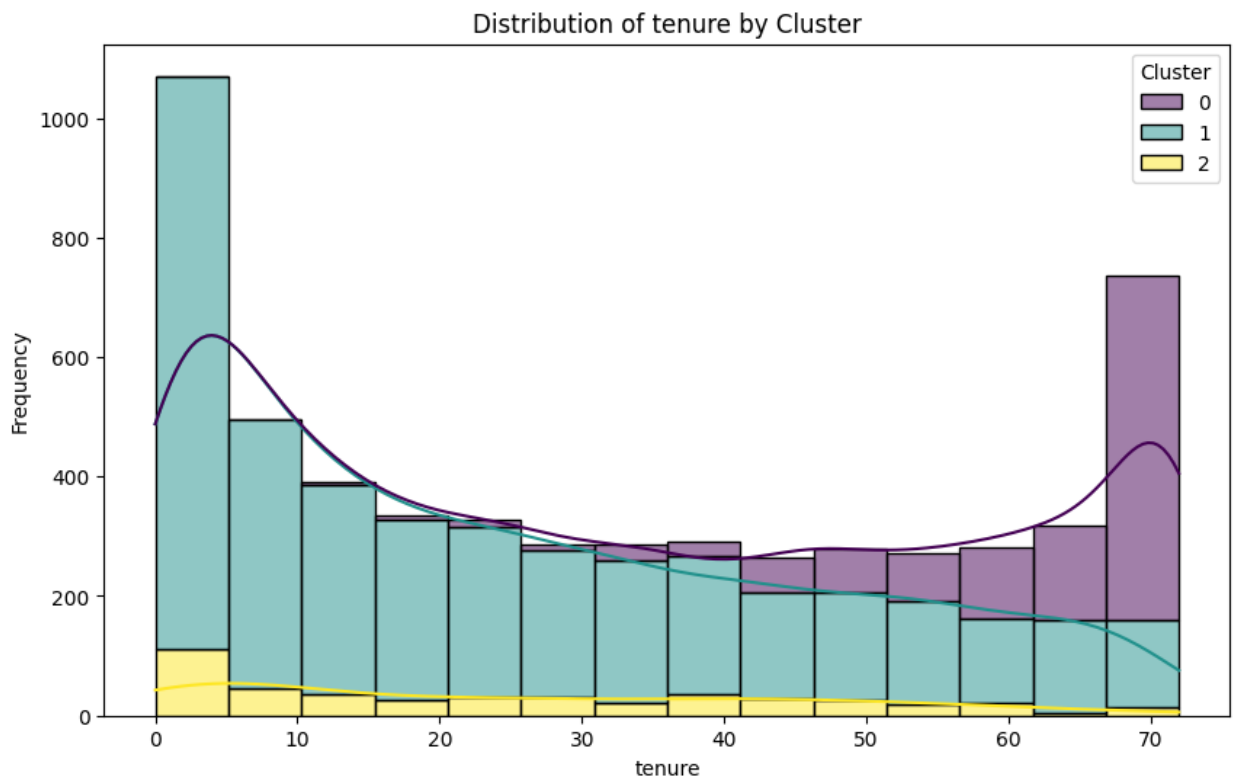


Figure 10 Distribution of Tenure by Cluster

3.2. Factors Identified through Predictive Analytics:

3.2.1. Key Predictors of Churn:

- **Tenure:** The ANN model identified tenure as a critical factor in predicting churn. Customers with shorter tenures were found to have a higher likelihood of churning, reinforcing the observations from the segmentation analysis.

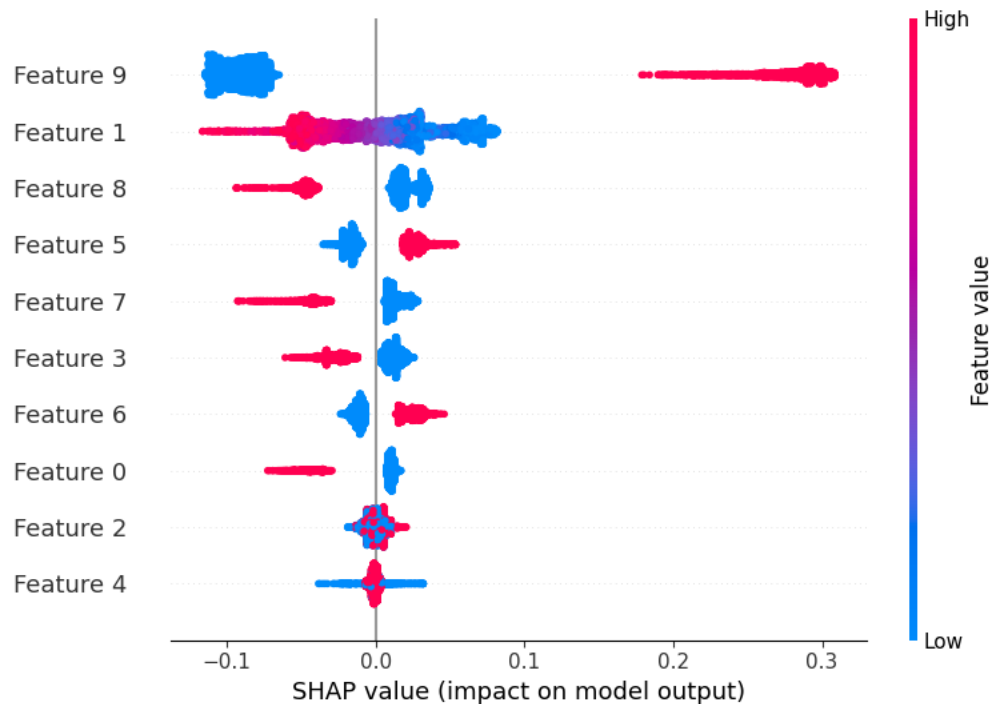
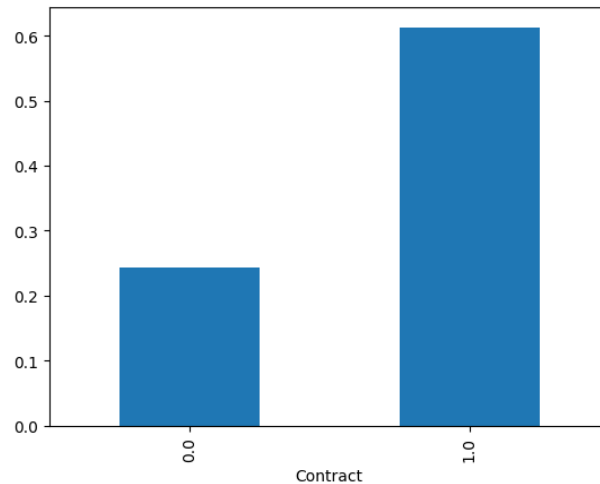


Figure 11 Feature Importance using SHAP Value

- **Internet Service Type:** The type of internet service, particularly whether the customer uses fiber optic, was another significant predictor. Fiber optic users had a higher churn probability, possibly due to unmet expectations for high-speed internet service.
- **Contract Length:** The model also highlighted contract length as a key predictor. Customers on shorter, more flexible contracts (e.g., month-to-month) were more prone to churn compared to those on longer-term contracts, where the commitment is higher, and the churn rate is lower.

3.2.2. Retention Indicators:

- **High Tenure:** Longer tenure was identified as a strong retention indicator, as customers who have been with the company for a long time are less likely to churn.



- **Bundled Services:** Customers subscribing to multiple services or family-oriented plans were more likely to be retained, as the convenience and cost savings associated with these bundles create a strong incentive to stay.
- **Positive Service Experience:** Although not directly measurable by a single feature, the model's predictions suggest that overall satisfaction with services plays a significant role in retaining customers, particularly in the high-tenure cluster.

4. Recommendations for Targeted Retention Strategies

4.1. Cluster-Specific Strategies:

Cluster 0: Moderate Tenure High Churn Fiber Optic Users

Customers in Cluster 0, characterized by moderate tenure and high usage of fiber optic services, are at a critical juncture in their relationship with the company. Their moderate tenure suggests that they have been with the company long enough to form opinions about the service but not long enough to be highly loyal. The high churn rate in this cluster indicates dissatisfaction, possibly due to unmet expectations regarding the performance of fiber optic services or a perceived lack of value compared to competitors.

To reduce churn among this group, the company should implement several targeted strategies:

1. **Tailored Retention Plans:** Offer personalized retention plans that address the specific needs and concerns of fiber optic users. For example, introducing service bundles that

include enhanced customer support, higher data caps, or premium content could add value and reduce the likelihood of churn.

2. **Service Quality Improvements:** Given the importance of service performance for this group, it's crucial to ensure that the fiber optic service meets or exceeds customer expectations. Investing in infrastructure upgrades, minimizing downtime, and enhancing service reliability can directly address the concerns that may be driving customers away.
3. **Proactive Customer Service:** Initiate proactive outreach to customers in this cluster before they reach the decision to churn. For example, customer service representatives could contact moderate tenure customers to offer free service evaluations, troubleshoot existing issues, or provide special offers. This proactive approach shows the customer that the company values their business and is committed to addressing their needs.
4. **Loyalty Programs:** Introduce loyalty programs that reward moderate tenure customers for staying with the company. These programs could include discounts for long-term contracts, rewards for referring friends, or points that can be redeemed for service upgrades. By recognizing and rewarding loyalty early, the company can increase customer retention.

Cluster 1: High Tenure Low Churn Family-Oriented Customers

Cluster 1 consists of customers with high tenure who are likely using family-oriented services. This group exhibits a strong level of loyalty, reflected in their low churn rates. However, maintaining this loyalty requires ongoing effort, particularly in ensuring that the services provided continue to meet the evolving needs of these customers.

Strategies to maintain and further enhance loyalty among this group include:

1. **Family-Friendly Promotions:** Continue to offer and expand family-friendly promotions that provide tangible value to this segment. Examples include discounts on family plans, free upgrades for adding more lines or services, and special deals on premium family content (e.g., parental controls, children's programming).
2. **Customer Experience Enhancement:** Focus on enhancing the overall customer experience, especially for long-term customers who may have higher expectations. This

could involve offering more personalized customer service, creating exclusive offers for long-tenure customers, or providing early access to new services and features.

3. **Engagement Programs:** Develop engagement programs that keep these customers actively involved with the brand. For instance, offering educational content on maximizing the value of their services, hosting family-oriented events, or creating a community forum where customers can share experiences and tips can deepen their connection with the brand.
4. **Recognition and Rewards:** Implement recognition programs that acknowledge the loyalty of high-tenure customers. Personalized thank-you notes, anniversary gifts, or exclusive access to new products can make these customers feel valued and appreciated, further solidifying their loyalty.

Cluster 2: Low Tenure High Churn Minimal Service Users

Cluster 2 includes customers with low tenure who use minimal services. These customers have a high churn rate, likely due to a lack of engagement with the company's offerings and possibly the ease with which they can switch to a competitor offering a similar or more attractive service.

To reduce churn in this segment, the company should consider the following interventions:

1. **Incentives to Explore Additional Services:** Offer incentives that encourage these customers to explore and use additional services. For example, providing free trials, discounts on bundles, or promotions that offer additional services at a reduced rate could increase their engagement with the company.
2. **Enhanced Onboarding Process:** Improve the onboarding process for new customers in this segment to ensure they fully understand the value of the services they have subscribed to. This could include personalized onboarding sessions, interactive tutorials, or welcome packages that highlight additional services they might find useful.
3. **Early Engagement Strategies:** Implement strategies that engage customers early in their relationship with the company. Regular follow-ups in the first few months, targeted communication that emphasizes the benefits of long-term engagement, and personalized offers can help convert low-tenure customers into loyal, long-term clients.

4. **Customer Feedback Loops:** Create feedback loops that allow these customers to express their needs and concerns early. By actively seeking out and responding to feedback from low tenure customers, the company can address issues before they lead to churn.

4.2. Overall Retention Strategy:

4.2.1. Predictive Analytics Integration:

The insights gained from the Artificial Neural Network (ANN) model should be seamlessly integrated into the company's broader retention strategies. By identifying high-risk customers early, the company can take proactive measures to retain them. The predictive model can be used to trigger automated alerts when a customer is likely to churn, allowing customer service teams to intervene with targeted offers or personalized communication.

Additionally, the model's predictions can help the company prioritize its retention efforts, focusing resources on the customers who are most likely to leave. This approach ensures that the company's retention strategies are both efficient and effective, targeting the right customers with the right interventions at the right time.

4.2.2. Personalization and Proactive Engagement:

Personalization is key to effective customer retention. Using the insights from the predictive model, the company can tailor its communication and offers to meet the specific needs of each customer segment. For example, high-risk customers might receive personalized discounts or service upgrades, while low-risk customers could be engaged with loyalty rewards and recognition programs.

Proactive engagement is also critical. Rather than waiting for customers to signal their dissatisfaction, the company should use the predictive model to identify potential churners and engage them before they make the decision to leave. This could involve automated email campaigns, personalized customer service outreach, or targeted marketing efforts designed to re-engage at-risk customers.

By integrating predictive analytics with personalized and proactive engagement strategies, the company can significantly reduce churn and build stronger, longer-lasting relationships with its customers.

4.2.3. Leverage Data-Driven Decision Making:

The deployment of predictive analytics in customer retention strategies allows for a more data-driven approach to decision-making. With the ANN model's insights, the company can transition from a reactive to a proactive retention strategy, where decisions are based on predictive indicators rather than waiting for churn to happen.

For example, the company can create a dashboard that tracks key metrics identified by the ANN model—such as tenure, service usage patterns, and contract types—providing real-time insights into the health of customer relationships. This dashboard could be used by customer service teams, marketing departments, and management to monitor churn risk across the entire customer base, ensuring that the company remains responsive and agile in its retention efforts.

Additionally, the integration of predictive analytics into decision-making processes allows for continuous learning and optimization. As more data is collected and the model is retrained, the company can refine its retention strategies based on the evolving needs and behaviors of its customer base.

4.2.4. Cross-Departmental Collaboration:

To maximize the impact of predictive analytics, it is crucial that insights from the ANN model are shared across departments within the company. The marketing team, customer service, product development, and even the sales department should all be aligned in their efforts to reduce churn and improve retention.

For instance, marketing campaigns can be tailored to target high-risk customers identified by the model, offering them promotions or incentives that are most likely to resonate based on their behavior patterns. Similarly, customer service teams can be alerted to focus their efforts on customers who are at high risk of churn, providing them with the necessary tools and information to offer personalized solutions.

By fostering cross-departmental collaboration, the company can ensure that every customer touchpoint is optimized for retention, from the initial marketing message to ongoing customer support and beyond.

4.2.5. Continuous Monitoring and Feedback Loops:

Predictive analytics is not a one-time solution but an ongoing process that requires continuous monitoring and refinement. The company should establish feedback loops to assess the effectiveness of its retention strategies and the accuracy of the ANN model's predictions.

For example, after implementing a targeted retention campaign based on the model's insights, the company should monitor the results closely to determine if the strategies were successful in reducing churn. This data can then be fed back into the model to improve its accuracy and predictive power.

Moreover, customer feedback should be actively solicited and incorporated into the model. Understanding why certain customers stay or leave, directly from the source, provides invaluable data that can help refine predictive models and retention strategies. This ensures that the company remains customer-focused and responsive to changing customer needs.

4.2.6. Personalization at Scale:

One of the most significant advantages of using predictive analytics is the ability to personalize retention efforts at scale. The ANN model enables the company to move beyond one-size-fits-all strategies, allowing for a more nuanced approach where each customer's individual needs and behaviors are considered.

For example, if the model identifies that a particular customer is likely to churn due to dissatisfaction with their current service bundle, the company can automatically generate a personalized offer that addresses this concern—such as a discounted upgrade or a service that better aligns with the customer's usage patterns.

By leveraging automation and AI, the company can deliver personalized experiences to thousands of customers simultaneously, ensuring that each interaction is relevant, timely, and impactful. This not only improves retention but also enhances customer satisfaction and loyalty, as customers feel understood and valued.

5. Documentation of Limitations and Proposed Solutions

5.1. Project Limitations:

1. Data Quality Issues: One of the primary challenges encountered during the project was the quality of the data available for analysis. The dataset had instances of missing values,

inconsistencies in data entry, and possible biases in the sample that could affect the accuracy of the model's predictions. Specifically, missing data points for critical features such as customer tenure or service usage required careful preprocessing and imputation, which may have introduced some level of uncertainty into the model.

2. Model Overfitting: During the training of the Artificial Neural Network (ANN) model, there were instances where the model exhibited signs of overfitting. This was particularly evident when the model performed exceptionally well on the training data but showed reduced accuracy on the validation set. Overfitting can result from the model being too complex, with too many parameters relative to the amount of data available, leading it to learn noise in the training data rather than generalizable patterns.

3. Limited Feature Set: The model was trained on a relatively limited set of features, which may not have captured all the complexities and nuances of customer behavior that influence churn. While the selected features provided significant insights, other potential predictors (e.g., customer interaction history, satisfaction scores, or external factors like market trends) were not available in the dataset, possibly limiting the model's overall effectiveness.

4. Lack of Real-Time Data: The analysis was conducted on a static dataset, which limits the applicability of the model in real-time scenarios. Customer behavior can change rapidly, and a model trained on historical data might not accurately predict future churn if customer preferences or external conditions shift significantly. The absence of real-time data integration poses a limitation in adapting the model to dynamic market conditions.

5. Interpretation of Complex Models: ANN models, while powerful, are often considered "black boxes" because of their complexity and the difficulty in interpreting how they make decisions. This can be a limitation when it comes to explaining the model's predictions to stakeholders or using its insights to inform broader strategic decisions. Although tools like SHAP values were used to improve interpretability, the complexity of the model may still pose challenges in practical applications.

5.2. Proposed Solutions:

1. Data Quality Improvement: To address data quality issues, future projects should prioritize data cleansing and validation during the initial stages. Implementing robust data collection

processes, automated validation checks, and regular audits can help ensure higher data accuracy. Additionally, acquiring supplementary data from alternative sources or conducting targeted data collection campaigns could fill in the gaps and enhance the model's reliability. This may also involve working closely with the data engineering team to establish better data governance practices.

2. Mitigating Overfitting: Several strategies can be employed to reduce the risk of overfitting in the ANN model:

- **Regularization Techniques:** Incorporating regularization methods such as L1, L2, or dropout layers can help prevent the model from becoming too complex and learning noise from the training data.
- **Cross-Validation:** Using k-fold cross-validation can provide a more robust evaluation of the model's performance, ensuring that it generalizes well to unseen data.
- **Simplifying the Model:** Reducing the number of hidden layers or neurons in the ANN can also mitigate overfitting by simplifying the model, making it less prone to learning noise.
- **Early Stopping:** Implementing early stopping based on the performance on a validation set can prevent the model from overtraining on the training data.

3. Expanding the Feature Set: To enhance the predictive power of the model, future efforts should focus on expanding the feature set to include more diverse and relevant data points. This could involve integrating customer satisfaction surveys, interaction logs, or external market data. Additionally, exploring feature engineering techniques to create new, informative features from existing data could improve the model's ability to capture complex customer behaviors.

4. Incorporating Real-Time Data: Integrating real-time data streams into the model could significantly enhance its applicability and accuracy in predicting churn. This would involve setting up real-time data pipelines, potentially using technologies like Apache Kafka or real-time databases, to feed the model with the latest customer interactions and behaviors. A real-time model could then be continuously updated and retrained to adapt to changing customer dynamics, providing more accurate and timely predictions.

5. Enhancing Model Interpretability: To address the challenge of interpreting complex ANN models, the use of advanced explainability tools like SHAP or LIME can be further explored. These tools provide detailed insights into how individual features contribute to model predictions, making it easier to communicate results to stakeholders. Alternatively, simpler models like decision trees or logistic regression could be used in tandem with the ANN to provide a more interpretable baseline, with the ANN serving as a more powerful but less transparent complement.

6. Continuous Model Evaluation and Tuning: Regularly monitoring and evaluating the model's performance after deployment is crucial to ensuring its continued effectiveness. Setting up automated monitoring systems that track key performance indicators (KPIs) such as accuracy, precision, and recall over time can help identify when the model's performance starts to degrade. This can trigger a retraining process or adjustments to the model architecture or features, ensuring that the model remains aligned with current customer behaviours and trends.