

INTRODUCTION:

In the competitive telecom industry, retaining customers is critical to sustaining revenue and growth. This project builds a machine learning model to predict customer churn and identify key drivers behind it. The workflow includes data aggregation via MySQL, EDA, feature engineering, and model building using Logistic Regression, Ridge, Lasso, and ElasticNet. PCA is used for dimensionality reduction, while SHAP explains model predictions. Finally, customers are segmented into actionable groups like At Risk, Loyal, and Dormant to guide retention strategies.

ABSTRACT:

Customer churn poses a major challenge in the telecom sector due to high competition and low switching costs. This project uses SQL and Python to aggregate and analyze customer data, applying machine learning models like Ridge, Lasso, and ElasticNet for churn prediction. Dimensionality reduction with PCA and explainability with SHAP enhance model insights. The outcome enables customer segmentation and targeted retention strategies to reduce churn and boost lifetime value.

Tools:

The project uses **Python** for data preprocessing, modeling, and visualization. **MySQL** is used to aggregate telecom usage data such as call duration, complaints, and recharges. **Pandas**, **NumPy**, and **scikit-learn** handle data manipulation and machine learning. **Matplotlib** and **Seaborn** are used for exploratory data visualization. **SHAP** is implemented for model explainability, and Machine learning for prediction and ide Google colab.

STEPS:

1. **Import and explore** the telecom customer dataset.
2. **Aggregate usage data** (call duration, complaints, recharges) using **MySQL**.
3. **Clean and preprocess** the data by handling missing values and encoding categorical features.
4. **Perform EDA** to identify patterns and trends related to customer churn.
5. **Apply feature engineering** to create new meaningful variables.
6. **Normalize** the data using **StandardScaler** for better model performance.
7. **Reduce dimensionality** with **PCA** and visualize data in 2D and 3D.
8. **Train classification models** (Logistic Regression, Ridge, Lasso, ElasticNet).
9. **Tune hyperparameters** using **GridSearchCV** and evaluate models with R^2 and MSE.
10. **Use SHAP for model explainability** and segment customers into **At Risk**, **Loyal**, and **Dormant** groups for actionable insights.

CONCLUSION:

This project successfully demonstrates how machine learning can be applied to predict customer churn in the telecom industry. By leveraging SQL, EDA, PCA, and various classification models, we identified key drivers of churn. SHAP analysis provided transparency into model decisions. Customer segmentation enabled targeted retention strategies. Overall, this data-driven approach supports proactive decision-making to reduce churn and enhance customer loyalty.

FINDINGS:

1. **21% of customers** were identified as likely to churn based on usage patterns and complaints.
2. Customers with **low call duration** and **infrequent recharges** had the **highest churn probability**.
3. **High complaint frequency** was a strong indicator of churn behavior.
4. The **ElasticNet model** achieved the **best performance** with an R^2 score of **0.82** on the test set.
5. **Top 5 features** influencing churn (via SHAP): complaint count, recharge frequency, call duration, total charges, and internet usage.
6. PCA helped reduce dimensions from **30+ features to 3**, preserving **90% of data variance**.
7. Customers were successfully segmented into:
 - **At Risk:** 18%
 - **Loyal:** 62%
 - **Dormant:** 20%
8. The model revealed that **prepaid users** are more likely to churn compared to postpaid users.
9. Customers with **multiple service issues in the last month** had a **3x higher churn risk**.
10. Targeted retention strategies can help save **up to 60%** of at-risk customers by early intervention.

FINAL RECOMMENDATIONS

1. **Target At-Risk Customers Early**
Use churn prediction scores to proactively engage customers showing signs of churn (e.g., low usage, high complaints).
2. **Improve Customer Service Quality**
Reduce churn by addressing complaint resolution times and offering loyalty rewards for frequent complainers.
3. **Offer Personalized Recharge Incentives**
Provide customized discounts or bonus plans to users with low recharge frequency or declining usage patterns.
4. **Monitor and Engage Dormant Users**
Re-engage inactive or dormant customers through special offers, surveys, and retention calls.
5. **Enhance Postpaid Plans**
Since prepaid users showed higher churn rates, consider bundling benefits or upselling to postpaid plans.
6. **Use Segmentation to Guide Strategy**
Develop different communication and service strategies for **At Risk**, **Loyal**, and **Dormant** segments.
7. **Continuously Update the Model**
Retrain the churn prediction model regularly with new data to maintain accuracy and relevance.