

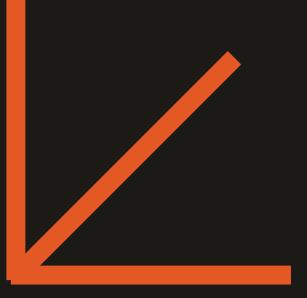


TELCO CHURN PREDICTION MODEL.

BY: NKATHA IMANENE



TABLE OF CONTENT



- 1**
TITLE
- 2**
**TABLE OF
CONTENT**
- 3**
OVERVIEW
- 4**
**BUSINESS &
DATA
UNDERSTANDING**
- 5**
**DATA
PREPARATION
AND
PREPROCESSING**
- 6**
**MODELING
APPROACH**
- 7**
**EVALUATION
METRICS**
- 8**
**MODEL
PERFORMANCE
RESULTS**
- 9**
**MODEL
INTERPRETATION
WITH SHAP**
- 10**
**RECOMMEND
ATIONS**



OVERVIEW

- Objective:
To develop a model that predicts customer churn and provides actionable insights to reduce churn rates.
- Key Goals:
 - 1.Understand customer behavior and factors influencing churn.
 - 2.Build predictive models to identify customers likely to churn.
 - 3.Provide recommendations to retain at-risk customers.



BUSINESS & DATA UNDERSTANDING

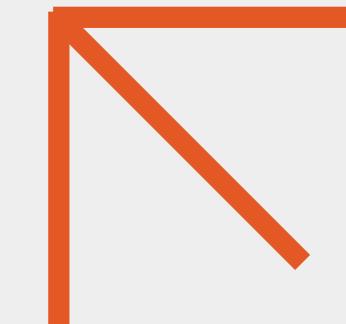
- Business Problem:

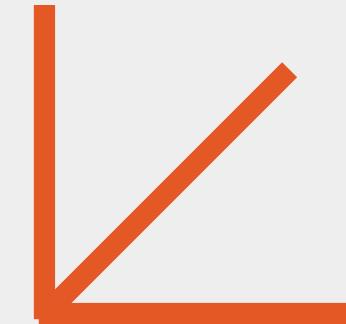
High customer churn rate is impacting revenue and growth. The goal is to identify customers likely to churn and understand the key drivers behind their decision.

- Data Used: Data includes customer demographics, account information, and service details.
- Key Variables:
 - Demographics: Gender, age, partner status, dependents.
 - Account Information: Tenure, contract type, payment method, charges.
 - Services: Phone service, internet service, online security, etc.



DATA PREPARATION AND PREPROCESSING

- Data Cleaning: Handled missing values, encoded categorical variables, and normalized numerical features.
 - Feature Engineering:
 - Created new features (e.g., combined service counts).
 - Applied SMOTE and undersampling to balance class distribution.
 - Splitting the Data: Split the dataset into 80% training and 20% testing data.
- 



MODELING APPROACH

- Models Used:
 - **Logistic Regression:** As a baseline model for interpretability.
 - **Decision Tree:** To capture non-linear relationships.
 - **Random Forest:** An ensemble method for improved accuracy and robustness.
- Hyperparameter Tuning: Performed grid search to fine-tune parameters for Random Forest to improve performance.



EVALUATION METRICS

- Metrics Used for Evaluation:
 - **Accuracy:** Overall correctness of the model.
 - **Precision:** Correctly identified churners out of all predicted churners.
 - **Recall:** Model's ability to capture all actual churners.
 - **F1-Score:** Balance between precision and recall.
- Why These Metrics?:
 - Helps assess model performance and its applicability to the business problem of customer churn.





MODEL PERFORMANCE RESULTS

- **Logistic Regression:**
 - Accuracy: 73%
 - Key Strength: High recall for capturing churners.
- **Decision Tree:**
 - Accuracy: 73%
 - Key Insight: Provides a balance between precision and recall.
- **Random Forest:**
 - Accuracy: 78%
 - Key Strength: Best overall performance, good trade-off between false positives and true churners.



MODEL INTERPRETATION WITH SHAP

- **Explainability:**
 - Used SHAP (SHapley Additive exPlanations) to interpret the impact of features on churn predictions.
- **Key Findings:**
 - Tenure, contract type, monthly charges, and tech support are significant predictors.
 - Contract type (month-to-month) has the highest positive correlation with churn.
- **Visualizations:** Summary plots showcasing the top features contributing to customer churn.



RECOMMENDATIONS

- **Retention Strategies:**
 - **Target At-Risk Customers:** Focus retention efforts on customers with month-to-month contracts and high monthly charges.
 - **Improve Customer Experience:** Enhance tech support and online security services.
 - **Incentivize Long-Term Contracts:** Offer discounts or benefits for customers switching to longer-term contracts.
- **Monitor Key Indicators:** Regularly analyze churn drivers and adjust strategies accordingly.



NKATHA
MAUREEN

