ABDELGHAFOR'S

HACKATHON

Names

Fares Alaaeldin   **Ahmed** Abozaid   **Abdalla** Ellaithy

Team Name

Blue Team

DATA PREPROCESSING

01

Missing Values

- Found 11 null values in TotalCharges.
- **Solution:**
  - Imputed missing values using the nearest neighbor based on Euclidean distance.
  - Categorical columns encoded using Label Encoding.

02

Imbalanced Data

- **Churn Distribution:**
  - Churn = 0: 5174
  - Churn = 1: 1869
- Addressed with SMOTE and Tomek Links.

FEATURE ENGINEERING

Techniques
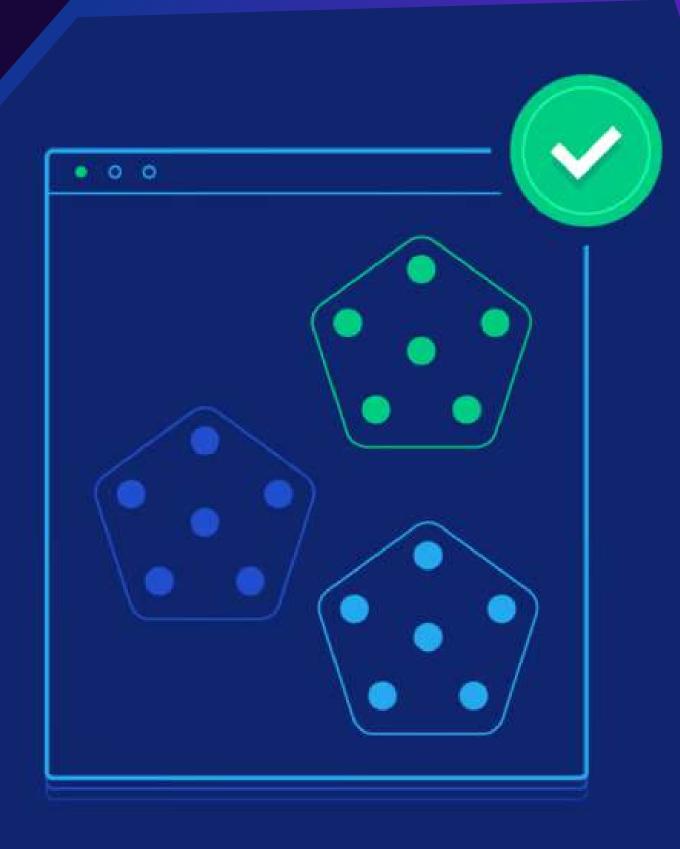
01 **Tenure Transformation:** Converted into time ranges (e.g., 0-1 year, 1-2 years).

02 **Charges per Month:** Created a new feature ChargesPerMonth = TotalCharges / Tenure.

03 **Interaction Features:**
- Contract_PaymentInteraction: to capture the relationship between contract type and payment method.

04 **Number of Services:** Aggregated services to create NumServices.

SUPERVISED LEARNING MODEL

MODEL PIPELINE

- **Models Tried:**
  - Random Forest
  - XGBoost
  - Logistic Regression
  - Decision Tree

- **Model Evaluation:**
  - Optuna Tuning: Accuracy: 86.6%, ROC-AUC: 93.8%
  - Grid Search: Accuracy: 86.7%, ROC-AUC: 93.9%
  - XGBoost: Accuracy: 85.8%, ROC-AUC: 94.1%

- **Chosen Model:** Random Forest
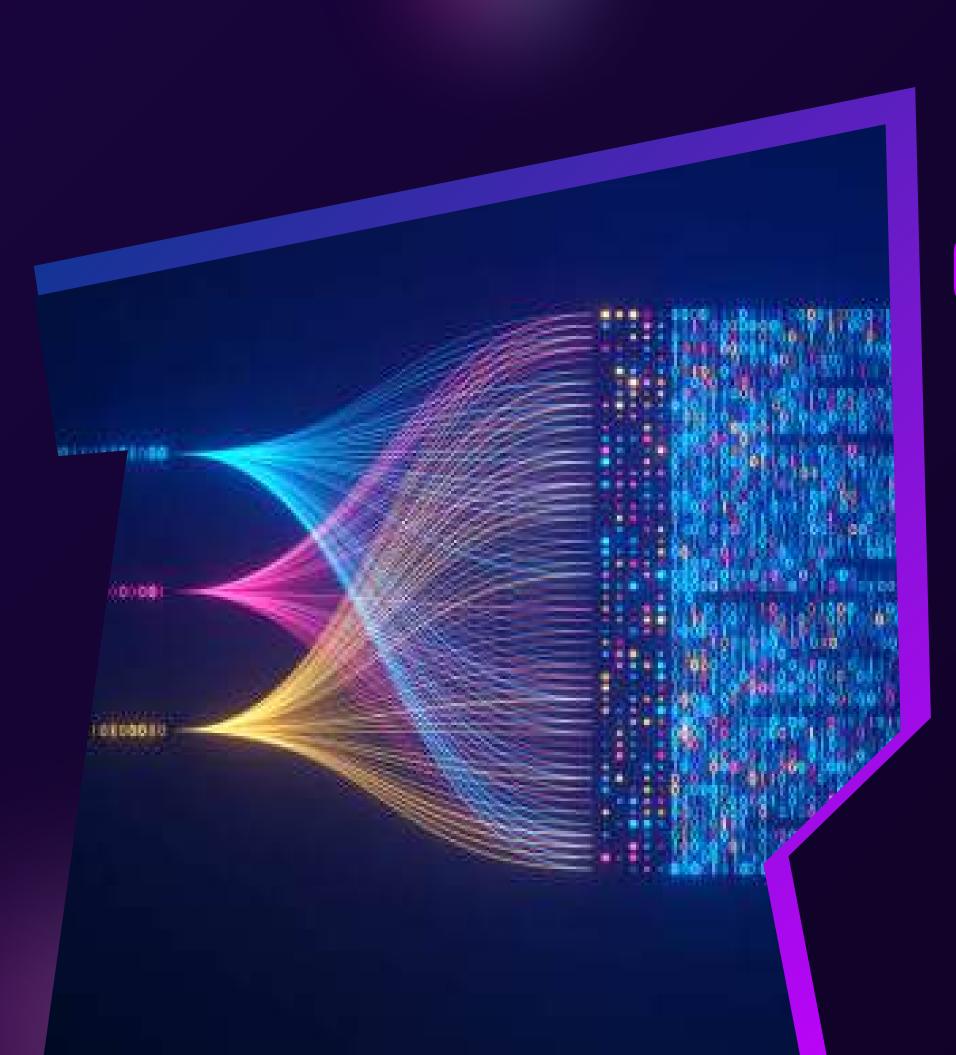
UNSUPERVISED LEARNING MODEL

Clustering

01 **K-Means:** Optimal clusters found with n=4 (using elbow and silhouette scores).

02 **Hierarchical Clustering:** Applied using linkage methods, visualized with dendrograms.

DATA VISUALIZATION

Visual Insights

01 **Churn Distribution:** Visualized imbalance.

02 **Feature Relationships:** Visualized key features like tenure and TotalCharges.

03 **Model Performance:** Plotted ROC-AUC and accuracy comparisons.

04 **Clustering Visualization:** Visualized customer segments using K-means and hierarchical clustering.

OVERALL ANALYSIS & INSIGHTS

KEY FINDINGS

01 **Churn Insights:** Gender had no significant impact on churn.

02 **Customer Segmentation:** Identified high-value customer segments for retention strategies.

03 **Actionable Recommendations:**
- Focus on long-tenure customers for upselling.
- Improve retention strategies for new customers.

Churn Reduction: Insights can help the company develop targeted interventions.

Customer Retention: Clustering insights can aid in customized offers for different customer segments.

BUSINESS IMPLICATIONS

ADVANCED TECHNIQUES

01

Advanced Feature Engineering

Interaction features, number of services.

02

Advanced Clustering

Hierarchical clustering beyond K-means.

03

Advanced Visualization

Used Plotly for interactive insights.

ACKNOWLEDGEMENTS

CONCLUSION

- **Improvement in accuracy:** Before feature engineering and SMOTE, model accuracy was between 70-81%. Post-improvements, accuracy increased to 86.7%.

- **Model chosen:** Random Forest for its performance.

- **Unsupervised learning:** K-means with optimal clusters identified.

Q&A

THANK YOU

Thank You