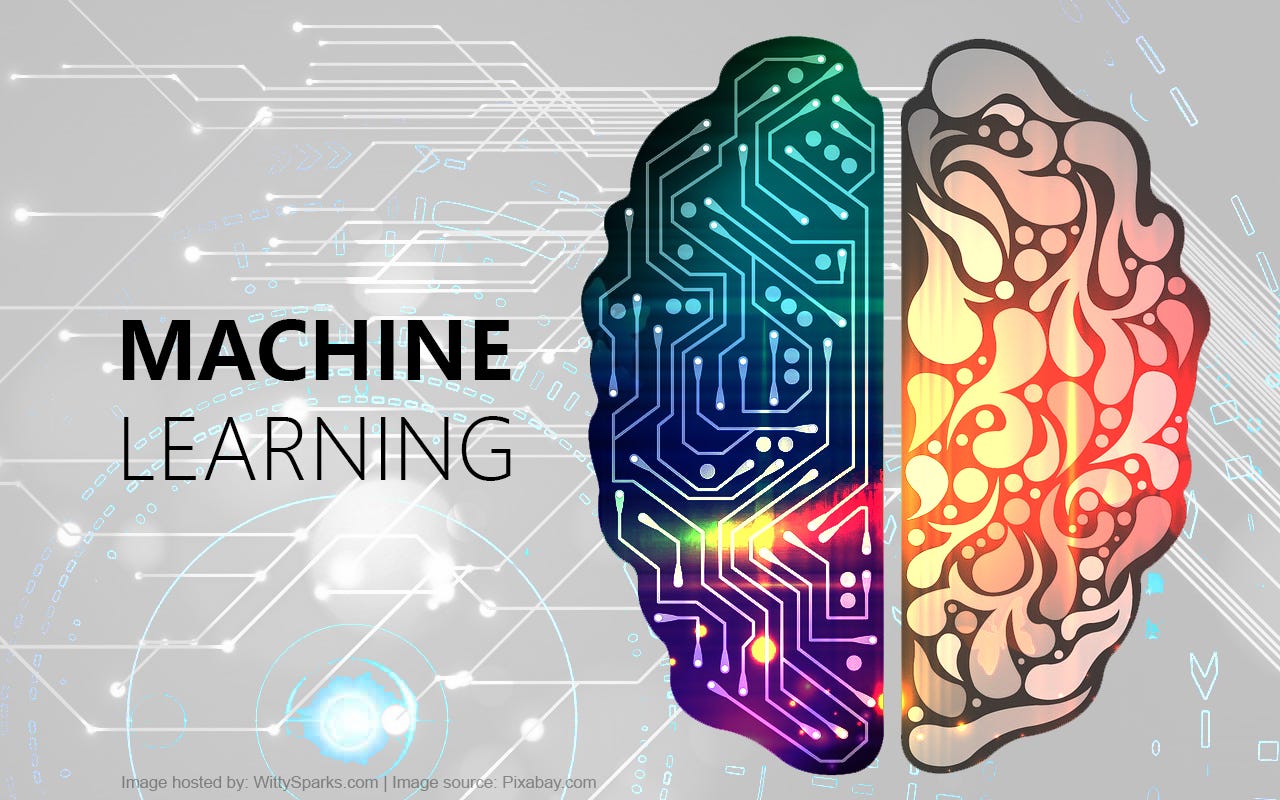
**CUSTOMER CHURN PREDICTION**

**MACHINE LEARNING**

**Angelite Arendse**

**Thembi Ngobeni**

**Zirong Luo**

**Armand Snyman**

Table of Contents

[**PROBLEM STATEMENT** 3](#_Toc196339574)

[**HYPOTHESIS GENERATION** 3](#_Toc196339575)

[**PROJECT OVERVIEW** 3](#_Toc196339576)

[**DATASET SUMMARY** 3](#_Toc196339577)

[**PREPROCESSING STEPS** 4](#_Toc196339578)

[**PREPROCESSING STEPS** 4](#_Toc196339579)

[**WEB APPLICATION WITH DASH** 4](#_Toc196339580)

[**DEPLOYMENT** 4](#_Toc196339581)

[**CONTINUOUS IMPROVEMENT** 4](#_Toc196339582)

[**REFLECTION AND FUTURE WORK** 5](#_Toc196339583)

[**PROJECT LINKS** 5](#_Toc196339584)

[**TEAM & CONTRIBUTIONS** 5](#_Toc196339585)

## **PROBLEM STATEMENT**

Telecommunications companies face significant revenue loss due to customer churn — the phenomenon where users discontinue their service subscriptions. Predicting customer churn in advance can empower businesses to proactively engage at-risk customers with retention strategies. The challenge lies in accurately identifying which customers are likely to leave, based on historical and behavioural data. This project aims to build a predictive machine learning model using the Telco Customer Churn dataset to determine the likelihood of a customer churning. The solution also includes a user-friendly web interface for real-time predictions.

## **HYPOTHESIS GENERATION**

 **Null Hypothesis (H₀):**  
There is no relationship between a customer’s attributes (e.g., contract type, tenure, monthly charges) and their likelihood to churn.

 **Alternative Hypothesis (H₁):**  
A customer's attributes such as contract type, tenure, monthly charges, and total charges are statistically significant predictors of their likelihood to churn.

## **PROJECT OVERVIEW**

The goal of this project is to predict whether a customer of a telecommunications company is likely to churn (i.e., stop using the service). This insight enables the business to proactively retain customers and reduce revenue loss. The project leverages the Telco Customer Churn dataset from Kaggle and uses a machine learning pipeline combined with a user-friendly web application built in Dash. The model is deployed and accessible via a web browser.

## **DATASET SUMMARY**

* **Source:** Kaggle - Telco Customer Churn
* **Records:** ~7043
* **Features:** 21 columns including demographics, services used, tenure, billing, and churn status
* **Target Variable:** Churn (Yes/No)

**Key features used:**

* Tenure,
* MonthlyCharges,
* TotalCharges (numeric)
* Contract (categorical, one-hot encoded)

## **PREPROCESSING STEPS**

1. Converted TotalCharges from object to numeric and imputed missing values
2. Dropped unnecessary columns like customerID
3. Applied one-hot encoding to categorical columns
4. Scaled numeric columns using StandardScaler
5. Split data into training and test sets (80/20)

## **PREPROCESSING STEPS**

* **Algorithm:** Random Forest Classifier
* **Evaluation Metrics:** Accuracy, Confusion Matrix, Classification Report
* **Performance:**
  + Accuracy: ~80% (depends on dataset split)
  + Balanced handling of churn and non-churn cases

**Files saved using joblib:**

* churn\_model.pkl
* scaler.pkl
* feature\_names.pkl

## **WEB APPLICATION WITH DASH**

A web application was built using Plotly Dash with the following features:

* Inputs: Tenure, MonthlyCharges, TotalCharges, Contract type
* Button to trigger churn prediction
* Real-time output indicating whether the customer will churn or not

The backend loads the model and scaler, processes user inputs, and displays a live prediction.

## **DEPLOYMENT**

The app is deployed using **Render**:

* Files included: app.py, model files, requirements.txt, Procfile
* The service is hosted online and available to users through a browser

# **CONTINUOUS IMPROVEMENT**

* Feedback was gathered from peers and instructors
* Changes implemented based on feedback:
  + Improved UI layout
  + Added default input values for usability
  + Ensured input format consistency with the training dataset

# **REFLECTION AND FUTURE WORK**

Through this project, the team gained valuable experience in machine learning, deployment, and UI/UX design.

**Planned enhancements:**

* Expand Dash interface to include more input features
* Implement login functionality and session management
* Add visual analytics (e.g., churn probability graphs)
* Integrate real-time data input and monitoring
* Explore ensemble methods or AutoML for improved accuracy

# **PROJECT LINKS**

1. **GitHub Repo:** <https://github.com/AngeliteArendse/Customer_Churn_prediction.git>
2. **Live Web App:** *[Insert your Render link here]*

# **TEAM & CONTRIBUTIONS**

* Data preprocessing & model training: Angelite Arendse, Zirong Luo, Thembi Ngobeni
* Web application development (Dash): Zirong Luo, Armand Snyman
* Deployment & documentation: Armand Snyman, Angelite Arendse