

Mini Project: Customer Churn Prediction with ANN

Problem Statement

This mini project involves predicting whether a customer will churn (leave the company) or not based on demographic, service usage, and contract-related features. The dataset contains information such as gender, contract type, monthly charges, tenure, and internet services. Students are required to perform Exploratory Data Analysis (EDA), preprocessing, build Artificial Neural Network (ANN) models of varying complexities, and finally deploy the best model using Streamlit.

The goal is to understand how deep learning can be applied to classification problems and how models can be deployed for real-world usage.

Dataset Link

 Telco Customer Churn Dataset (Kaggle) —

<https://www.kaggle.com/datasets/mosapabdelghany/telcom-customer-churn-dataset>

Guidelines for Students

1. Data Understanding

- Explore the dataset structure (Telco-Customer-Churn.csv).
- Identify feature types:
 - **Numerical:** tenure, MonthlyCharges, TotalCharges
 - **Categorical:** gender, Partner, Dependents, PhoneService, InternetService, Contract, PaymentMethod, etc.
- Target column: **Churn** (Yes/No).

2. Data Exploration (EDA)

- Plot churn distribution (class imbalance check).
- Analyze churn percentage by contract type and payment method.
- Boxplots of tenure vs churn, MonthlyCharges vs churn.
- Correlation heatmap for numerical features.
- Compare churn rates across internet service types.

3. Preprocessing

- Encode categorical variables using OneHotEncoding.
- Scale numerical features using StandardScaler.
- Convert target variable **Churn** into binary (Yes=1, No=0).
- Split dataset into training, validation, and test sets (70/15/15).

4. Model Building

You will build and compare **three ANN models** of different complexities:

Model 1 (Simple ANN)

- Dense(20, relu) → Dense(1, sigmoid)
- Loss: Binary Crossentropy
- Optimizer: Adam

Model 2 (Intermediate ANN)

- Dense(32, relu) → Dropout(0.1) → Dense(16, relu) → Dense(1, sigmoid)
- Dropout prevents overfitting.
- EarlyStopping to stop when validation loss stops improving.

Model 3 (Advanced ANN)

- Dense(64, relu) → Dropout(0.2) → Dense(32, relu) → Dropout(0.2) → Dense(1, sigmoid)
- Larger architecture with stronger regularization.

5. Model Evaluation

- Evaluate models using: Accuracy, Precision, Recall, F1-score, ROC-AUC.
- Plot training vs validation loss curves.
- Compare performance of all three models.
- Best model achieved ~78.8% test accuracy.

✓ Saved best model as `best_ann_model1.h5` and preprocessor as `preprocessor.pkl`.

6. Deployment with Streamlit

Streamlit App Features

- User-friendly web form to collect customer details (gender, contract, monthly charges, tenure, etc.).
- Preprocess input using saved encoder & scaler.
- Load trained ANN model to predict churn probability.

- Display prediction as:
 - “**Customer likely to churn**”
 - “**Customer likely to stay**”

Hosting

- App hosted on **Streamlit Cloud**, accessible via public link.
- Demonstrates end-to-end ML project deployment.

7. Project Tasks

Basic Level

1. Display dataset shape and first 5 rows.
2. Check for missing values.
3. Plot churn distribution.
4. Compare churn rate by gender.
5. Analyze churn by contract type.

Intermediate Level

1. Encode categorical variables and scale numerical ones.
2. Train baseline ANN model (Model 1).
3. Report accuracy and F1-score for baseline model.
4. Plot training vs validation loss for baseline model.

Advanced Level

1. Train Model 2 and Model 3 with Dropout and EarlyStopping.
2. Compare accuracy, precision, recall, and AUC across all models.
3. Save the best performing ANN and preprocessing pipeline.
4. Build a Streamlit app for customer churn prediction.
5. Deploy app on Streamlit Cloud and share link.

Expected Outcomes

- **Basic:** Students will learn dataset exploration and visualization.
- **Intermediate:** Students will understand preprocessing and ANN training for classification.
- **Advanced:** Students will learn model comparison, optimization with regularization, and deployment on Streamlit Cloud.

By the end of this project, students will demonstrate the complete pipeline:

EDA → Preprocessing → ANN Models → Model Selection → Deployment → Evaluation & Interpretation.