**Customer Churn Prediction and Sales Dashboard**

**1. Introduction**

Customer churn prediction is a crucial task for businesses to identify customers who are likely to leave a service or product. By applying Machine Learning (ML) and Deep Learning (DL) models, organizations can proactively take steps to reduce churn and improve customer retention.

The main objective of this project is to build and evaluate ML/DL models to predict churn accurately.

**2. Dataset Overview**

* **Features**: Customer demographics, usage patterns, product/service category, etc.
* **Target**: churn\_flag (1 = churned, 0 = not churned).
* **Data Preprocessing Steps**:
  + Handling missing values.
  + Encoding categorical features.
  + Feature scaling/normalization.
  + Train-test split for supervised learning.

**3. Models Implemented**

**📌 Supervised Machine Learning Models**

| **Model** | **Accuracy** |
| --- | --- |
| Logistic Regression | **1.00** |
| Decision Tree | **1.00** |
| Random Forest | **1.00** |
| XGBoost | 0.67 |
| LightGBM | 0.67 |

➡️ Logistic Regression, Decision Tree, and Random Forest performed extremely well with **100% accuracy**.  
➡️ XGBoost and LightGBM showed **moderate performance (67%)**.

**📌 Deep Learning Models**

| **Model** | **Accuracy** |
| --- | --- |
| ANN | 0.33 |
| RNN | 0.67 |

➡️ ANN showed **low accuracy (33%)**.  
➡️ RNN performed better (67%) but still lower than tree-based ML models.

**📌 Unsupervised Learning (Exploratory)**

Unsupervised techniques such as **K-Means Clustering, DBSCAN, Hierarchical Clustering** were explored to segment customers, but they were not used for direct churn prediction (since they don’t use labels). They helped in identifying **hidden customer segments**.

**4. Model Comparison (Graph)**

📊 The following bar chart visualizes supervised ML model accuracies:

import matplotlib.pyplot as plt

models = ["Logistic Regression", "Decision Tree", "Random Forest", "XGBoost", "LightGBM"]

accuracies = [1, 1, 1, 0.67, 0.67]

plt.figure(figsize=(8,5))

plt.bar(models, accuracies)

plt.title("Supervised Model Accuracies - Churn Prediction")

plt.ylabel("Accuracy")

plt.xticks(rotation=30)

plt.show()

**5. Key Findings**

* **Tree-based models (Decision Tree, Random Forest)** achieved the best performance.
* **Logistic Regression** also performed excellently, likely due to a clear separation in churn vs non-churn data.
* **Boosting models (XGBoost, LightGBM)** underperformed, possibly due to small dataset size or overfitting issues.
* **Deep learning (ANN, RNN)** did not perform well compared to classical ML models, showing that simpler models are more effective for this dataset.

**6. Recommendations**

* Use **Random Forest or Logistic Regression** for production deployment.
* Further improve boosting models by **hyperparameter tuning** and **feature engineering**.
* Explore **ensemble stacking** to combine multiple models.
* Regularly retrain models with **new customer data** to ensure accuracy.

**7. Conclusion**

This churn prediction project demonstrates that **classical ML models (especially Random Forest and Logistic Regression)** outperform deep learning and boosting models on the given dataset. The insights can help businesses proactively target at-risk customers and reduce churn.

**8. Screenshot**

