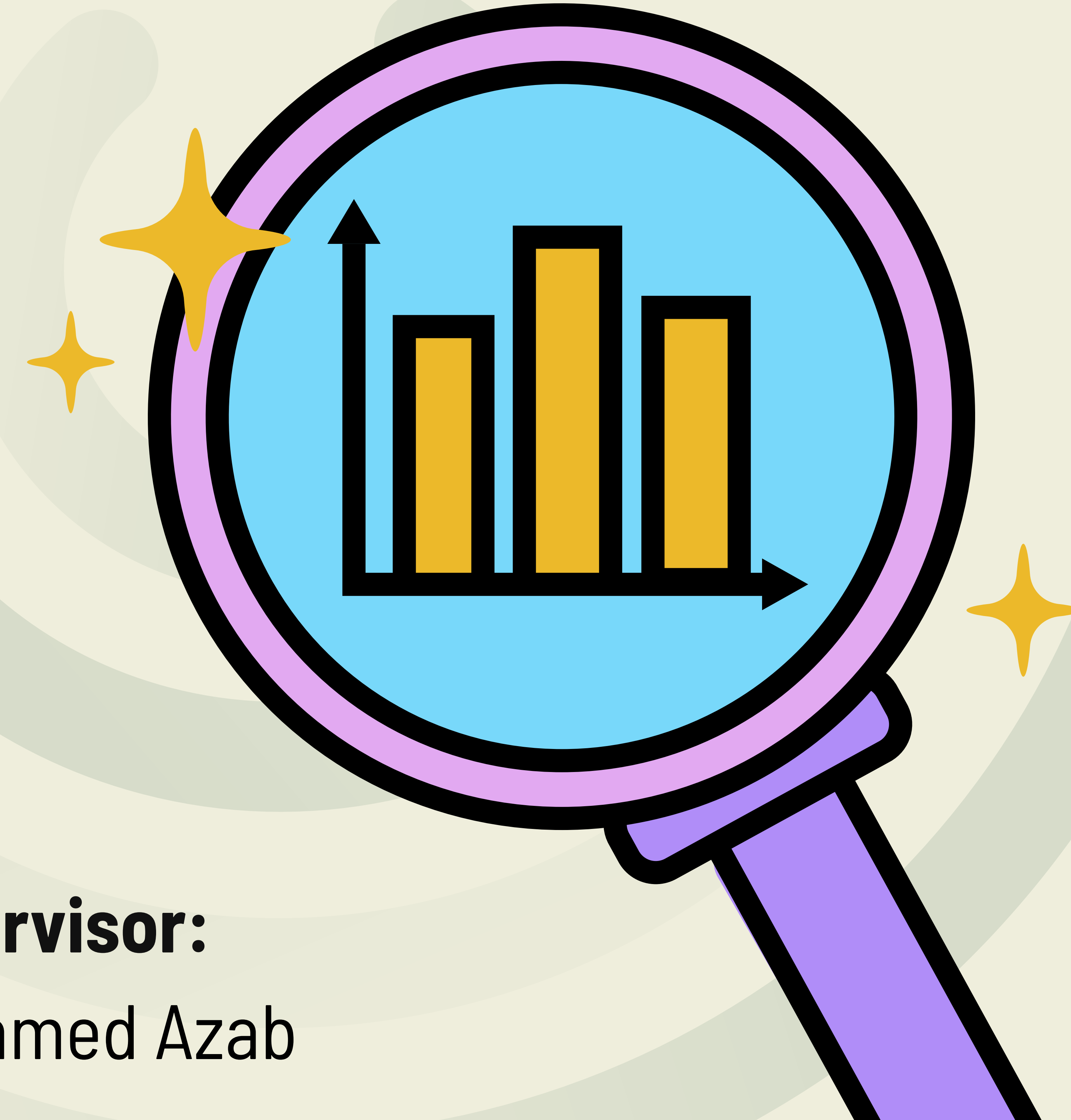
رواد مصر الرقمية

DEPI R3 Graduation Project
Data Engineering Track
Group 34

Customer Churn



Presented By

Yakoot Shaker
Nancy Nabil
Nada Ahmed
Abdelrahman Ayman
Ahmed Rashad

Supervisor:
Prof. Ahmed Azab

رواد مصر الرقمية

Problem Statement

Why is this problem important?

Customer churn is a critical issue because acquiring new customers is significantly more expensive than retaining existing ones. Companies lose revenue every time a customer leaves, so predicting churn early helps prevent financial loss.

What losses occur due to churn?

High churn reduces Customer Lifetime Value (CLV), increases marketing and acquisition costs, and weakens the competitive position of the business.

Dataset Overview

رواد مصر الرقمية

Main Features Used:

- tenure
- usage
- complaints
- plan_type
- monthly_charges
- support_calls
- Target Variable: churn_status

Key Characteristics from the Documentation:

- The dataset has a clear class imbalance (most customers are Non-Churn).
- Strong relationships exist between churn and:
- → complaints
- → low tenure
- → Basic plan_type

رواد مصر الرقمية

Data Preprocessing

Based on the preprocessing pipeline described:

- Mean imputation for usage
- One-Hot Encoding for plan_type
- StandardScaler for all numerical features
- 80/20 train-test split

Exploratory Data Analysis (EDA)

Main insights reported:

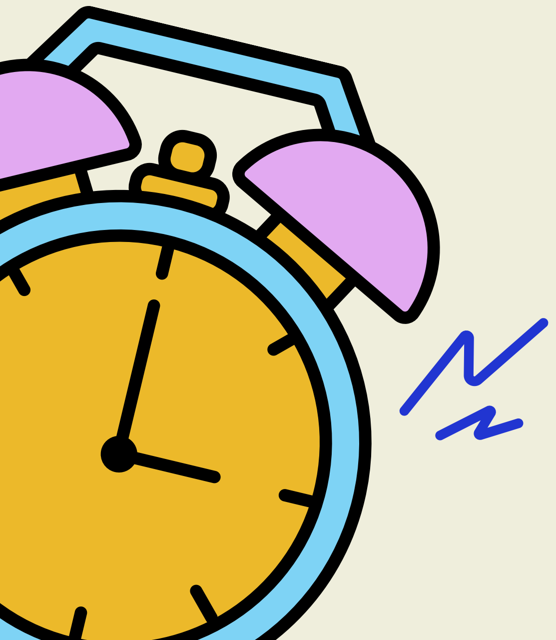
- **Class imbalance:** majority are "Not Churned"
- **Higher churn** was observed among customers who:
- → Filed many complaints
- → Had low tenure (especially < 6 months)
- → Used the Basic plan
- Complaints showed a **strong direct relationship** with churn likelihood
- Tenure had a **bimodal distribution**, indicating different churn behaviors

Modeling

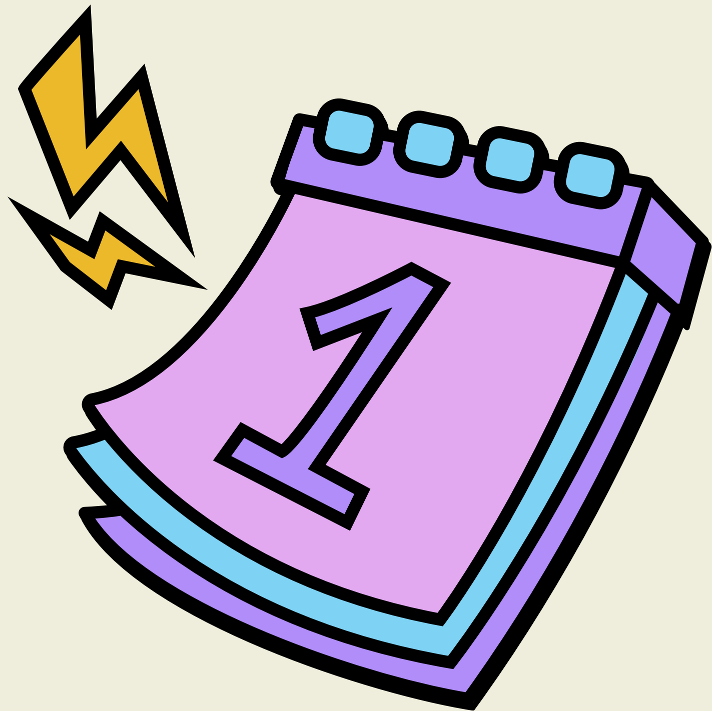
The following models were compared:

Model	**Purpose**
Logistic Regression	Baseline, interpretable
Random Forest	Non-linear baseline
XGBoost	Advanced model — final chosen model

Why XGBoost was selected:



- Best overall performance
- Handles class imbalance effectively
- Captures complex feature interactions
- Provides strong interpretability through feature importance



Performance Metrics

Metric	**Logistic Regression**	**Random Forest**	**XGBoost (Final)**
Accuracy	0.82	0.87	**0.89**
Precision	0.65	0.78	**0.81**
Recall	0.7	0.75	**0.8**
F1-Score	0.67	0.76	**0.8**
ROC-AUC	0.84	0.88	**0.91**

Final Model Performance

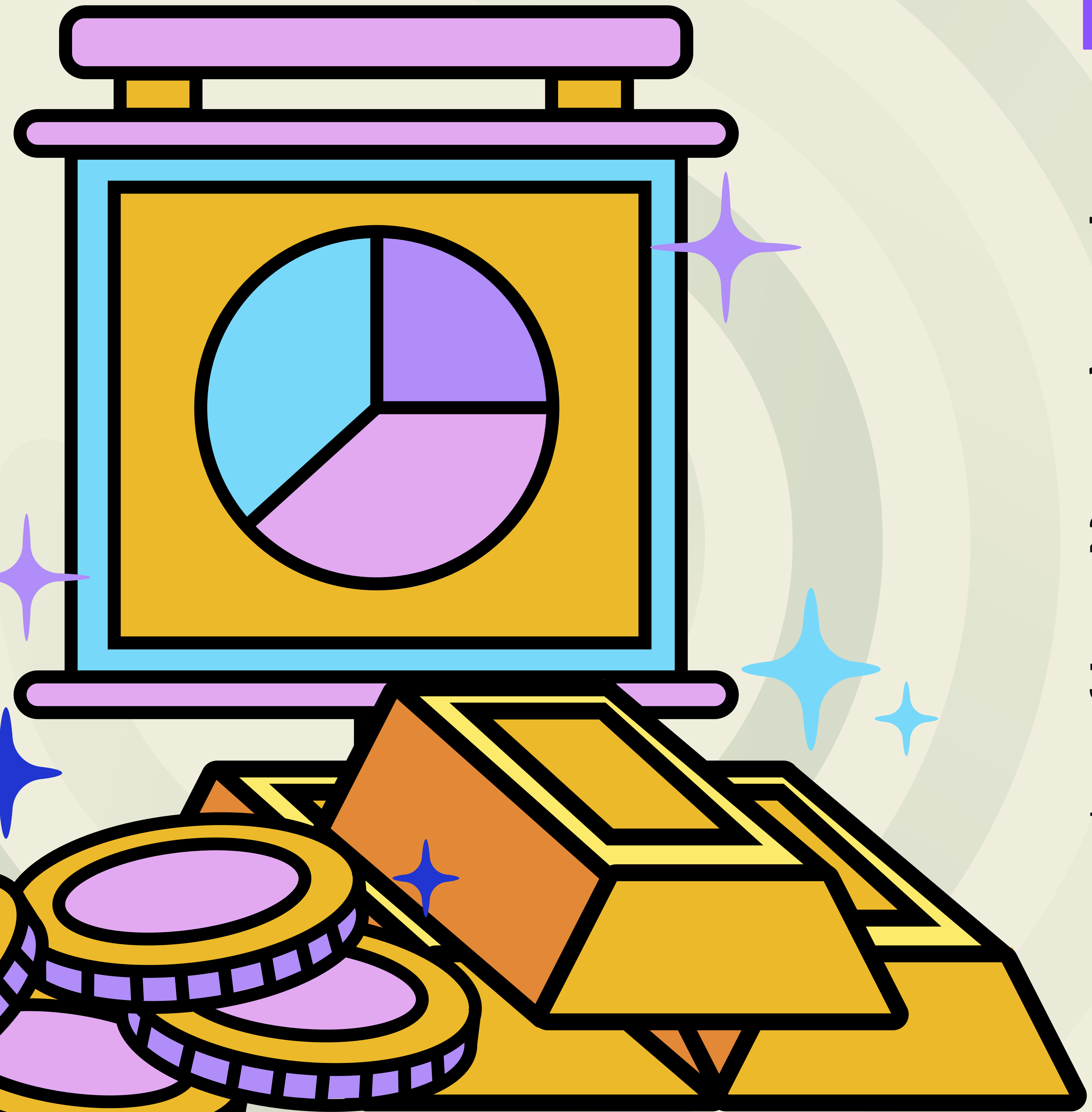
Final Chosen Model: XGBoost

Key Performance Metrics:

- Accuracy: 0.89
- Recall (churn): 0.80
- ROC-AUC: 0.91

Include a **Confusion Matrix** in the slide visually.

The model showed a strong balance between precision and recall,



Feature Importance

Top features influencing churn:

1. complaints – the strongest predictor

2.tenure – customers with short tenure churn more

3.plan_type – Basic plan users had the highest churn

These insights help explain why the model makes certain predictions and support business decisions.

Conclusion

- The team built a complete end-to-end churn prediction system.
- The final model achieved **high predictive performance** (Accuracy 0.89).
- **The system can help companies:**

→ Identify at-risk customers early
→ Reduce churn
→ Improve customer satisfaction

- **With more time, the project could be extended with:**

→ A richer UI
→ Additional behavioral features
→ More automation

Future Work

رواد مصر الرقمية

- Integrate the model with a **live customer database**
- Automate **periodic model retraining**
- Build a **churn analytics dashboard** inside the Streamlit app
- Enhance the product with more real-time features

Use Case 1: Ministry of Education

Student Dropout Prevention

- **Input:** Attendance records, grades, socio-economic data.
- **Prediction:** Identify students at high risk of dropping out months in advance
- **Intervention:** AI suggests personalized support programs or social worker visits to address root causes (e.g., financial aid).

رواد مصر الرقمية

Use Case 2: Ministry of Health

Patient Follow-up & Adherence

- **Problem:** Patients with chronic diseases stopping treatment.
- **Prediction:** Forecast patients likely to miss appointments based on refill rates.
- **Intervention:** Automated, empathetic AI calls/messages to remind patients, reducing burden on public hospitals.

Use Case 3: Ministry of Social Solidarity

Takaful & Karama Support

- **Problem:** Identifying families vulnerable to economic shocks.
- **Prediction:** Analyze economic indicators to predict household financial distress.
- **Intervention:** Proactive allocation of support funds before a crisis occurs.

رواد مصر الرقمية

رواد مصر الرقمية

Thank You