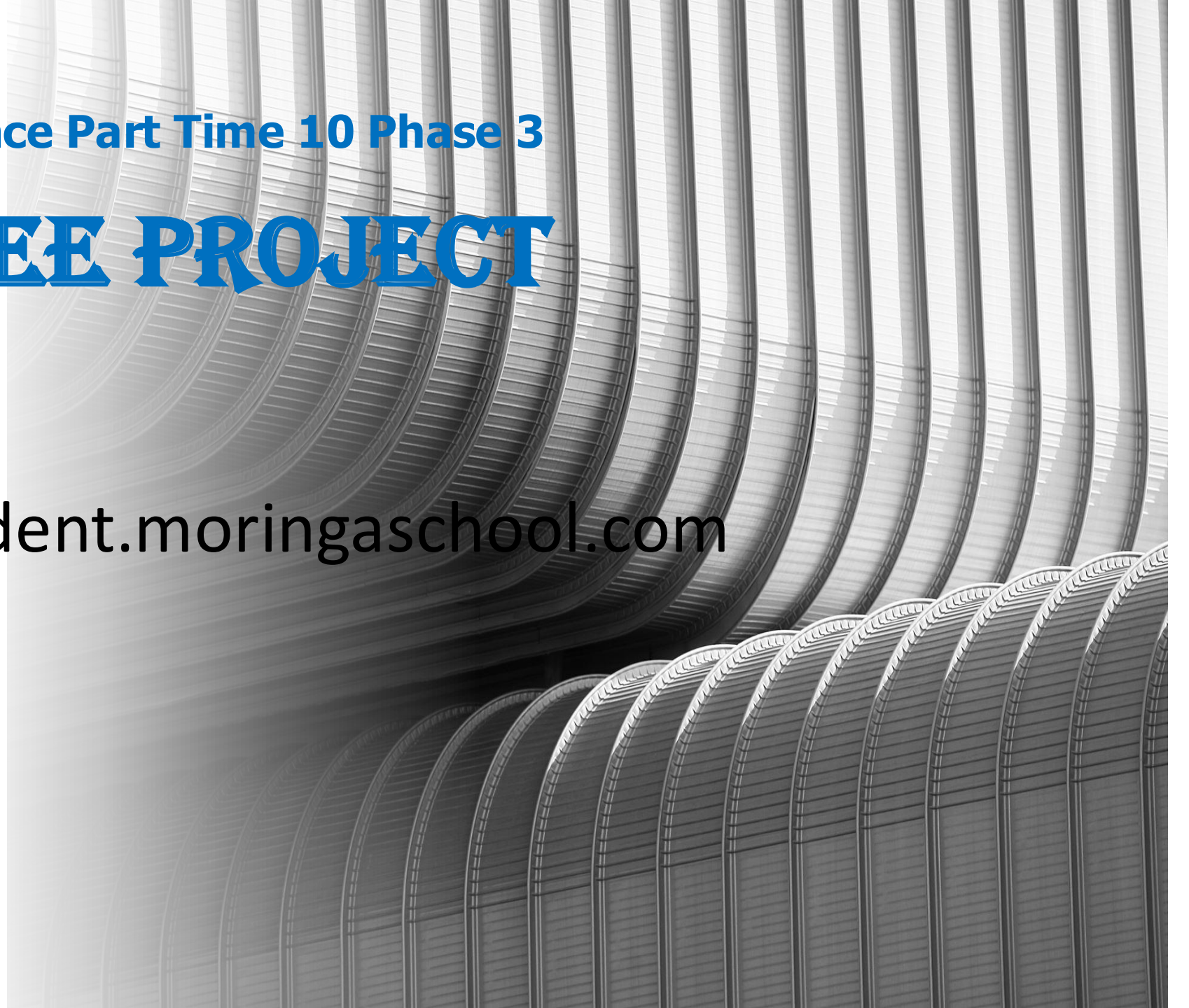
DSF-PT10P3: Data Science Part Time 10 Phase 3

PHASE THREE PROJECT

JABES KOTIENO

jabes.kotieno@student.moringaschool.com



Business Overview and Problem Definition

Project Title:

- *Dialing into Retention - Predicting Customer Churn at SyriaTel Using Machine Learning*

Objective:

- Support SyriaTel, a leading telecommunications provider, in reducing customer churn.
- Churn directly impacts recurring revenue and increases customer acquisition costs.
- Identifying high-risk customers will allow SyriaTel to target retention efforts more effectively and optimize resources.

DATA DESCRIPTIONS AND UNDERSTANDING

Dataset Overview:

- The data used contains key customer information relevant to predicting churn, such as customer usage patterns, payment history, and demographic details.
- *The CSV file-bigml_59c28831336c6604c800002a.csv (applied as bigml_59.csv) is available at https://www.kaggle.com/datasets/becksddf/churn-in-telecoms-dataset.*

- **Exploratory Data Analysis (EDA)**:
  - Identified issues such as missing values, outliers, and class imbalance.
  - Conducted univariate and bivariate analyses to uncover relationships and patterns in the data.
  - Focused on understanding customer behavior that correlates with churn.

FEATURE SELECTION AND ENGINEERING

Feature Selection:

Features were selected based on correlation analysis and importance scores.

Derived variables created, such as "total charges per call" and flags for high usage.

Dimensionality reduction techniques like feature pruning were applied to improve model efficiency.

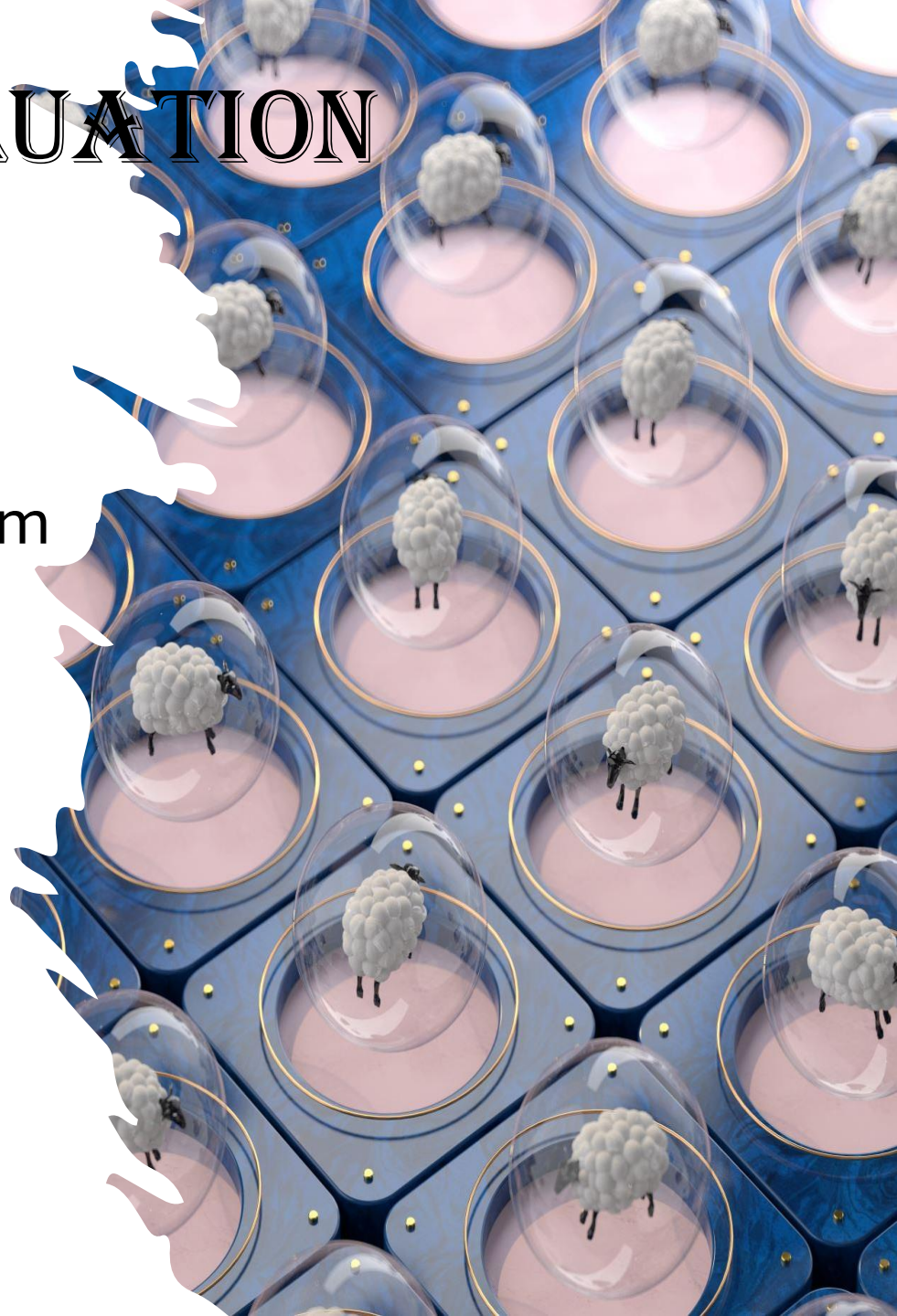
Model Building and Evaluation

Modeling Approach:

- Started with baseline models (logistic regression, decision trees).
- Trained advanced models, including Random Forest and Gradient Boosting.

Evaluation Metrics:

- Accuracy, precision, recall, F1-score, and ROC-AUC were used to assess the models' performance



Interpretation and Recommendations

Key Insights:

- Feature importance analysis revealed key factors influencing customer churn.
- Visualized decision boundaries and model outcomes to improve transparency.

Business Recommendations:

- Target high-risk customer segments for retention actions.
- Integrate churn prediction model into SyriaTel's CRM or customer service platform.
- Improve data collection processes for future model iterations.