# Grant Acquisition Data Science Use Case

Chinonso Amamchukwu

# Introduction: Overview

- The faculty members at the Lagos Business School conduct research focused on solving business challenges in the African environment and the world at large. Part of conducting impact-driven research is to secure grant funding from various grant institutions to solve business and societal problems.

- Faculty members apply for funding opportunities in different research areas of the funders' interest. The application fails or succeeds.

- Ideal question: Can I build a model that can predict the outcome of each faculty application for grant funding? Does the applicant's academic status influence his/her chances of acquiring the grant?

# Data Science Use Case using Dorard's ML Canvas

Let's represent the ML canvas as in the diagram:

# Problem Statement

- **Priority:** To increase research revenue at LBS by increasing grant acquisition frequency.

- Operational terms:

- Acquisition frequency: how often do researchers secure or acquire grant funds?

- - Grant fund: financial assistance given to researchers to support their research projects for a particular purpose.

# Value Proposition

- Increase grant acquisition frequency, thus increasing research revenue. This is a strategic objective. +12%.
- Detect trending research topics that would provide more chances to secure grant funds.
- Boosting application success rate.

# Data Sources
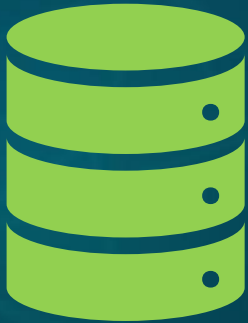


- Historical and unstructured data in the LBS research archive. Xlsx and doc.

- Grantor's websites for areas of interest.

# Data Collection

Historical data of grant application records

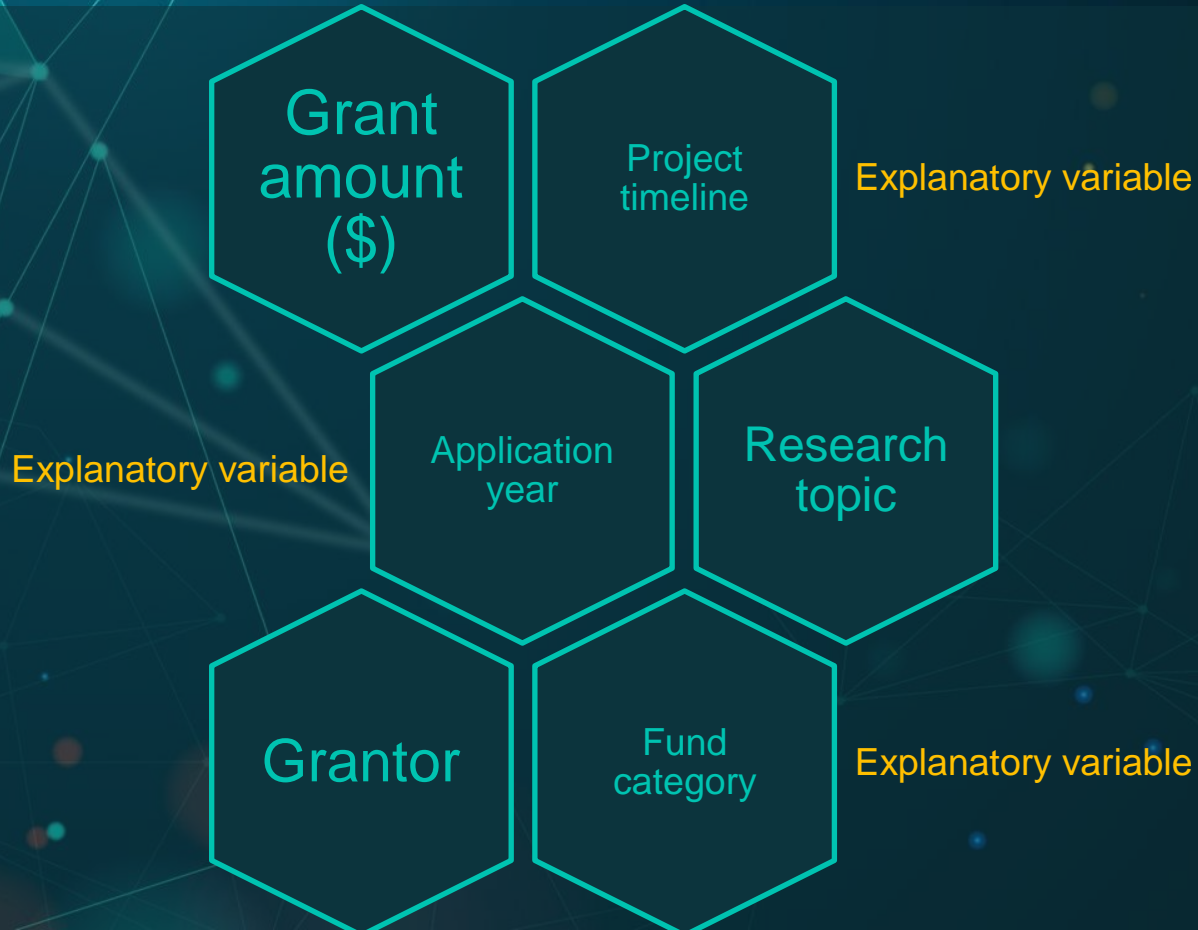To find out faculty research areas of interest being targeted for the grant application.

Research international organisations' agenda, since it directs research. For example, SDG 2030 by the United Nations

**Archival Data & Unstructured data**

**Polls & Survey**

**International bodies' agendas**

# Features [Engineering]

Grant amount ($)

Project timeline

Explanatory variable

Explanatory variable

Application year

Research topic

Grantor

Fund category

Explanatory variable

# Dataset sample

Features → | ← Target →

| # | Grantor | Research area | Application year | Grant amount | Project timeline | Fund category | Grant | |
|---|---------|---------------|-----------------|--------------|------------------|---------------|-------|---|
| 1 | Sunref Nigeria & Access bank | Energy | 2022 | 317,623 | 2 | Partnership | 1 | |
| 2 | Ford Foundation and Innoson Chukwuma Social Impact Chair and Fellowship | Sustainability | 2022 | 1,000,000 | 5 | Long-term | 1 | |
| 3 | Templeton Charity Foundation | Virtual Reality | 2021 | 234,000 | 3 | Long-term | 1 | |

# Machine Learning Task

❑ **Predictive TASK:** Predict the outcome of grant application: WHICH GRANT APPLICATION IS LIKELY TO BE ACCEPTED?

❑ **INPUT:** Faculty Research

❑ **OUTPUT:** "Likely" or "Unlikely".

❑ **APPROACH:** Binary Classification Problem. Supervised Learning from the research features (attributes) or explanatory variables.

# Implementation

- Which platform is best for the implementation of the task?

- OPTIONS:

  - Cloud: AMAZON MACHINE LEARNING –*Batch processing* for a group of observations.

    - This is costly, even for *Real-time processing for single observation interaction.*

  - Personal Computer. Core i7 processor, 8th generation, SSD 256, 2.4 GHz.

# DECISION

- **At the end of each quarter:** Randomly select 80% of the grant applications. Use it as test data.

- Why Quarterly?

- Grantors' make sponsorship calls at intervals depending on their area of interest. As such, data generated from grant applications does not have the high velocity as big data.

# Building the Model

- Consider the Machine Learning Workflow:
- From raw data > Extract FEATURES (after Data Preprocessing) > Split dataset (train and test data) > Train data > Train Model > Use test data to EVALUATE the model performance.

- METRICS:

    - Accuracy and Confusion Matrix. Observe the false positives (T1) and false negatives (T2). Observe Overfitting (Model good at training, bad at testing).
- IF the model is not good enough, I retrain the model by adding or removing some features via:

    - Dimensionality reduction (columns): using correlation, Hyper-parameter tuning or ensemble method.

# Make Prediction

## Likely (1) or Unlikely (0)

EACH QUARTER, FEATURE ALL CURRENT GRANT APPLICATIONS AND MAKE PREDICTIONS ABOUT THEM

# Impact Simulation/ Offline Evaluation

- The evaluation will be mostly based Offline, on a quarterly basis, from historical grant application data.

- Before targeting grant outcome, consider

    - Evaluating the new model's accuracy on predefined researcher profiles.

    - Simulate decisions made on last quarter's faculty application (using a model that was learned from the last two quarter's grant application outcomes).

# LIVE EVALUATION/MONITORING

❑ Amazon Machine Learning (if used) performs real-time processing and monitoring, on demand.

❑ Optionally,

❑ I can compare the grant acquisition rate and the application frequency between last quarter's dataset and the remaining set.

❑ Monitor secured grant in the targeted (application outcome) and its contribution to research revenue (KPI) based on the difference between failed applications and time taken to apply for the grant, or the ratio of the grant fund acquired to the financial ($xxx,000,000) target (KPI) of the research department in LBS.

# ML Canvas - DASHBOARD

## THE MACHINE LEARNING CANVAS

Designed for: **DAT 601 Assignment**  Designed by: **Chinonso Amamchukwu**

Date: **29 November 2022**  Iteration: **V_01.**

### PREDICTION TASK

ML Task: *Which grant application is likely to be accepted?*

- Input: Faculty Research
- Output: Likely or unlikely.
- Approach: Binary Classification Problem. A supervised learning approach.

### DECISIONS

Quarterly.

Randomly select 80% of the grant application and use it as test data.

### VALUE PROPOSITION

- Increase grant acquisition rate, thus increasing research revenue.
- Detect trending topics that would provide more chances to secure grant funds.
- Boost application success rate.

### DATA COLLECTION

- Collate historical data of grant applications in the record.
- Launch polls and surveys among researchers to find out the areas of interest being targeted in their respective grant applications.

### DATA SOURCES

- Historical and unstructured data in the LBS research archive. Xlsx and doc format.
- Grantors' websites for funds areas of interest.

### IMPACT SIMULATION

Offline/Impact simulation:

Before targeting grant outcome grant applications, evaluate the new model's accuracy on pre-defined researchers' profiles.

### MAKING PREDICTIONS

Each quarter, feature all current grant applications and make predictions about them.

### BUILDING MODELS

Each quarter, create a new model from the previous quarter's test data.

### FEATURES

The features are the following independent or explanatory variables: grant amount, project timeline, application year, research topic, grantor, and fund category.

### MONITORING

Monitor the secured grant and its contribution to the research revenue based on the ratio of the grant fund to the financial target (KPI) of the research department in LBS.

The End-to-End ML task might be a different use case. Data limitations.

# Next: Perform ML task

Daalu.
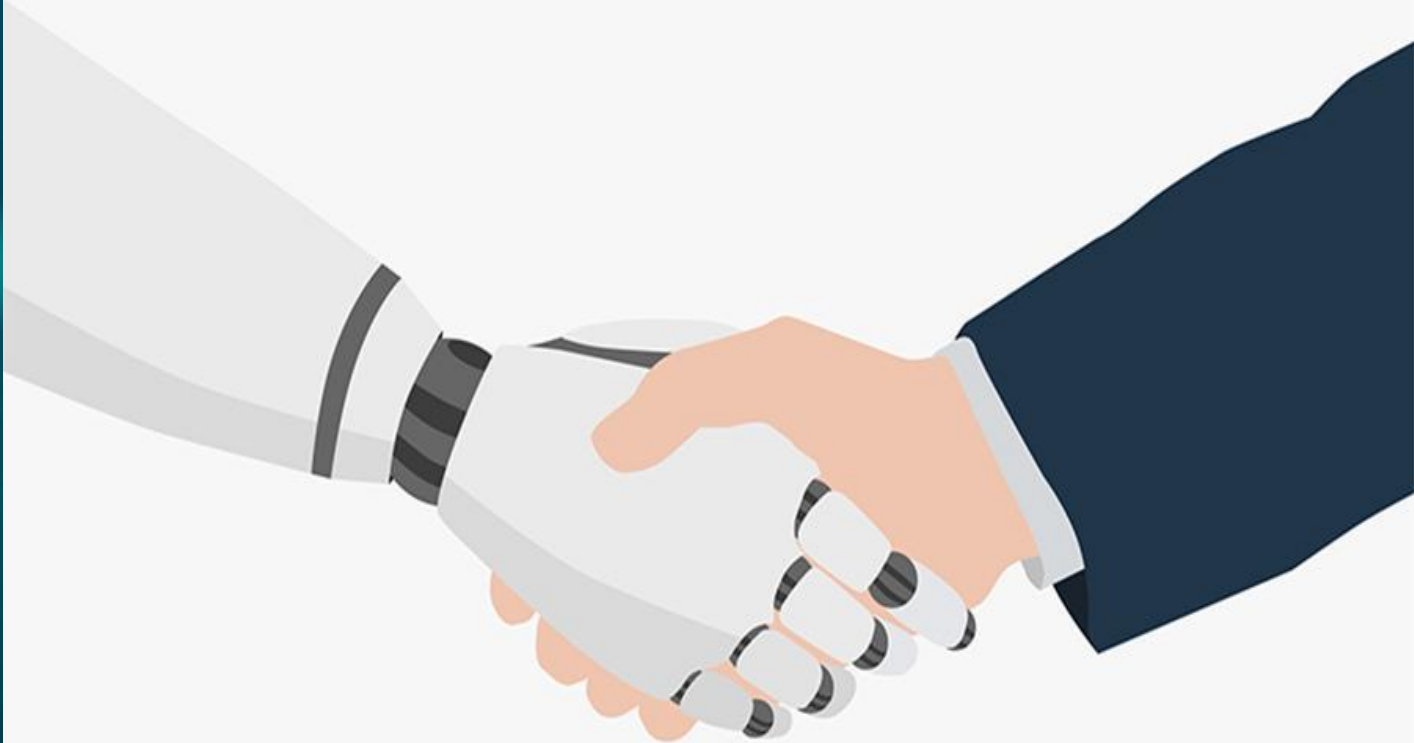
# LINK TO MY ZOOM PRESENTATION RECORDING

- GOOGLE DRIVE:

# [Click to view the video recording](#)