**ALY6110 70591 Data Management & Big Data SEC 08 Fall 2023 CPS**

**Module 4 Assignment — Final Project REPORT**

Logo

Description automatically generated

**NORTHEASTERN UNIVERSITY**

**College of Professional Studies, Boston, MA, 02215.**

**Submitted by**

Zihan Ma

ma.zihan1@northeastern.edu

**Instructor**

 Prof. Hema Seshadri

**Date**

12/01/2023

**Big Data Application in Bank Marketing Strategy**

**Dataset Overview:**

A dataset sourced from Kaggle containing 768,604 records will be used. It includes 17 variables such as age, job, marital status, and financial status, providing a comprehensive view of customers' interactions with bank marketing campaigns.

**Analytical Tools:**

Apache Spark will be employed for its efficient processing capabilities. The Random Forest algorithm will be used to predict term deposit subscription likelihood, and K-means clustering will identify distinct customer segments.

**Objective:**

The analysis aims to enhance marketing campaign effectiveness by identifying customers more inclined to subscribe to term deposits.

**Objective:**

* Use Spark to preprocess and manage the dataset.
* Apply Random Forest to find significant predictors for subscription likelihood.
* Perform K-means clustering to segment the customer base for targeted marketing.

**Expected Outcome:**

The project will provide insights to tailor marketing efforts, increasing the subscription rate for term deposits.

**Significance:**

The findings will contribute to optimizing bank marketing strategies, potentially leading to increased conversion rates and better resource allocation.