**LAPORAN PROYEK PENGOLAHAN DATA BESAR**

***Classification in Bank Marketing for Deposit Subscription Prediction Using Spark MLib***



**Disusun oleh:**

1. 12S17007– Ernike Nelsi Manurung
2. 12S17012 – Reza Oktovian Siregar
3. 12S17020 – Jovan Pioma Pakpahan
4. 12S17021 – Inggrit Purba

**PROGRAM STUDI SARJANA SISTEM INFORMASI**

**FAKULTAS INFORMATIKA DAN TEKNIK ELEKTRO**

**INSTITUT TEKNOLOGI DEL**

**APRIL 2020**

# DAFTAR ISI

[DAFTAR ISI i](#_gjdgxs)

[DAFTAR TABEL ii](#_30j0zll)

[DAFTAR GAMBAR iii](#_1fob9te)

[Bab I Pendahuluan 4](#_2et92p0)

# DAFTAR TABEL

No table of figures entries found.

# DAFTAR GAMBAR

# No table of figures entries found.

# Bab I Pendahuluan

1. **Latar Belakang**

Marketing Campaign adalah susunan strategi aktivitas pemasaran yang dimanfaatkan untuk melakukan promosi produk melalui berbagai media seperti televisi, radio, telepon seluler maupun platform online. Dalam bidang marketing terdapat 2 pendekatan periklanan dan promosi yakni mass marketing dan direct marketing [1]. Pada kasus didalam proyek ini akan dilakukan melalui media telepon seluler yang termasuk didalam direct marketing. Kegiatan marketing merupakan bagian penting didalam bisnis apapun, salah satunya adalah bank marketing. Kumpulan data Bank Marketing terkait dengan kampanye pemasaran langsung dari lembaga perbankan Portugis yang dimana kampanye tersebut didasarkan melalui panggilan telepon. Permasalahan yang sering terjadi pada kasus ini seperti ditemukannya lebih dari satu kontak ke klien yang sama diperlukan, untuk mengakses apakah produk (deposito berjangka/bank term deposit) akan ('ya') atau ('tidak') berlangganan. Didalam Bank Direct Marketing ditemukan aspek yang penting yakni bagaimana memperoleh akurasi yang tinggi didalam proses klasifikasi berdasarkan ketepatan informasi tertentu yang didapatkan dari pelanggan serta dianggap penting bagi pihak bank (Elsalamony, 2013).

Pihak Bank membutuhkan strategi untuk dapat mengelola data para nasabah yang berjumlah besar. Salah satu strategi yang dapat dilakukan untuk melakukan pengolahan informasi yang berasal dari data-data nasabah sebelumnya yakni dengan cara memprediksi. Melalui hasil prediksi tersebut dengan mengolah serta memanfaatkan data-data nasabah deposito yang sudah ada tersebut maka diharapkan akan menemukan informasi yang bersifat penting dan bernilai.

Marketing didalam banking service semakin kompleks diera Industri 4.0. Pada era ini sangat penting untuk menerapkan teknologi terbaru untuk dapat menyediakan pelayanan terhadap pengguna untuk menjadi lebih baik serta produktif. Salah satu teknologi yang mendukung hal tersebut adalah Apache Spark. Apache Spark merupakan perpaduan antara machine learning dan big data (unified engine) yang dapat melakukan pemrosesan untuk mengolah data yang besar. Apache Spark memiliki beberapa komponen yang mendukung proses pengklasifikasian, yaitu Spark MLib dimana Spark MLib menyediakan p

Pada proyek ini akan melakukan klasifikasi terhadap atribut data yang dapat dimanfaatkan untuk melakukan prediksi apakah klien akan berlangganan (ya / tidak) deposito berjangka sebagai strategi didalam pemasaran.

1. **Tujuan**

Adapun tujuan didalam melakukan penelitian pengolahan data besar ini yakni :

1. Memprediksi klien akan berlangganan deposito berjangka menggunakan metode klasifikasi pada kumpulan data Bank Marketing dengan menggunakan Spark Mlib API.
2. **Manfaat**

Manfaat dari pengerjaan proyek ini yaitu:

1. Bagi Mahasiswa

Proyek ini diharapkan nantinya dapat memberikan pemahaman dan keterampilan bagi tim proyek untuk mengimplementasikan teknik *machine learning (scalable machine learning)* pada data besar khususnya pada salah satu metode dari pengklasifikasian yaitu *Logistic Regression* dengan teknologi pengolahan data besar menggunakan Spark Mlib API untuk memproses dan menganalisis data besar.

1. Pihak Bank

Proyek ini diharapkan dapat digunakan oleh pihak bank untuk mengelola data para nasabah yang berjumlah besar agar dapat memprediksi apakah klien akan berlangganan (ya / tidak) deposito berjangka sebagai strategi di dalam pemasaran. Melalui hasil prediksi tersebut dengan mengelola serta memanfaatkan data-data nasabah deposito yang sudah ada tersebut maka diharapkan akan menemukan informasi yang bersifat penting dan bernilai.

1. **Ruang Lingkup**

Ruang lingkup dalam pengerjaan proyek ini adalah menggunakan salah satu metode dari Classification yaitu Logistic Regression. Dataset yang digunakan adalah data Bank Marketing yang diunduh dari UC Irvine Machine Learning Repository.

# Daftar Pustaka

|  |  |
| --- | --- |
| [1] | T. D. H. Phan Duy Hung and T. D. Tung, "Term Deposit Subscription Prediction Using Spark MLlib and ML Packages," 25-28 February 2019. [Online]. Available: https://doi.org/10.1145/3317614.3317618. [Accessed 26 April 2021]. |