**Memprediksi Peningkatan Pemasaran Bank**

**LAPORAN TECHNICAL REPORT**



**DISUSUN OLEH :**

**KELOMPOK 5**

**ANALICIA 22031554007**

**AHMAD RAFI SYAIFUDIN 22031554030**

**RIVA DIAN ARDIANSYAH 22031554043**

**S1 SAINS DATA**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

**UNIVERSITAS NEGERI SURABAYA**

**2023**

**DAFTAR ISI**

[**BAB I** 3](#_Toc154148330)

[**PENDAHULUAN** 3](#_Toc154148331)

[**1.1** **LATAR BELAKANG** 3](#_Toc154148332)

[**1.2** **Tujuan dan Manfaat** 3](#_Toc154148333)

[**BAB II** 4](#_Toc154148334)

[**METODE** 4](#_Toc154148335)

[**2.1** **RANDOM FOREST** 4](#_Toc154148336)

[**2.2** **DECISSION TREE** 4](#_Toc154148337)

[**2.3** **LOGISTIC REGRESSION** 4](#_Toc154148338)

[**BAB III** 5](#_Toc154148339)

[**HASIL DAN ANALISIS** 5](#_Toc154148340)

[**1.1** **USER MANUAL GUIDE** 5](#_Toc154148341)

[**1.2** **LISTING CODE** 6](#_Toc154148342)

[**1.3** **DIAGRAM ALIR** 11](#_Toc154148343)

[**DETAIL JOB PROJECT** 12](#_Toc154148344)

[**BAB IV** 13](#_Toc154148345)

[**KESIMPULAN** 13](#_Toc154148346)

[**REFERENSI** 14](#_Toc154148347)

## **BAB I**

## **PENDAHULUAN**

## **LATAR BELAKANG**

Project ini membahas berbagai jenis model algoritma untuk pengklasifikasian dengan menggunakan beberapa model, yaitu Random Forest, Decission Tree, dan Logistic Regression. Dengan tujuan untuk mengevaluasi dan menerapkan metode tersebut. Data yang digunakan berisi pemasaran suatu institusi bank. Project ini memprediksi terhadap suatu klien bank yang akan berlangganan deposito berjangka panjang (variabel y) berdasarkan atribut pemasaran oleh institusi bank.

## **Tujuan dan Manfaat**

Ada pun project ini dilakukan ialah bertujuan untuk melakukan klasifikasi terhadap klien bank yang terlihat memungkinkan untuk berlanganan deposito yang berjangka Panjang. Data bank yang digunakan masih tergolong inbalance sehingga dilakukannya resampling pada data. Kemudian data tersebut dilakukan tuning agar mendapat performa akurasi yang lebih baik dari sebelumnya.

Dalam hal ini dapat memberikan manfaat untuk meningkatkan efisiensi pemasaran sehingga bank dapat memberikan arahan pemasaran yang tepat. Manfaat dilakukannnya pengklasifikasian menggunakan ketiga algoritma tersebut ialah melihat prediksi yang dilakukan akurat untuk membandingkan dan mengevaluasi kinerja algoritma. Selain itu, juga dapat melihat bagaimana penggunaan informasi yang dikembangkan berdasarkan strategi pemasaran terhadap pelanggan maupun kebutuhan pelanggan tersebut.

## **BAB II**

## **METODE**

## **RANDOM FOREST**

Random Forest merupakan teknik dalam machine learning yang menggunakan metode pohon keputusan untuk mengevaluasi model sehingga melalui pohon keputusan ini digunakan untuk membuat keputusan yang memperlihatkan kondisi dari tiap cabang tersebut. Random forest sendiri memiliki set data latih yang berbeda untuk mengurangi overfitting dan meningkatkan ketahanan pada model itu sendiri.

Pada random forest dapat digunakan untuk menganalisis faktor dan strategi pemasaran bank terhadap pelanggan. Selain itu, dapat membantu dalam memprediksi risiko pelanggan terhadap deposito sehingga dari pihak bank dapat membuat keputusan yang lebih baik. Maka random forest ini dapat digunakan untuk memprediksi dan analisis terhadap peningkatan strategi pemasaran.

## **DECISSION TREE**

Decision tree merupakan machine learning yang juga digunakan untuk memprediksi atau mengklasifikasi data dengan mengambil keputusan dengan cara yang mirip dengan pohon kepututsan atau random forest. Decision tree mudah melakukan interpretasi dan visualisasi yang memudahkan dalam pengambilan keputusan, juga tidak memiliki tahap preprocessing yang rumit dan memiliki fitur otomatis dengan cara mengidentifikasi fitur yang penting dalam pembuatan keputusan.

Decision tree juga memiliki batasan yang dapat menyebabkan overfitting dan menjadi tidak stabil karena memiliki perubahan yang kecil didalam data. Meskipun demikian, machine learning ini tetap baik digunakan karena mudah dalam interpretasi dan pengklasifikasiannya yang relative sederhana.

## **LOGISTIC REGRESSION**

Logistic regression adalah suatu model statistik umum tentunya digunakan untuk menganalisis. Logistic regression digunakan untuk memodelkan suatu hubungan antara variable dependen yang memiliki dua nilai yang mungkin, seperti ya atau tidak, 0 atau 1, dan lain-lain. Logistic regression juga digunakan untuk memahami faktor-faktor yang signifikan terhadap strategi dalam pemasaran.

Logistic regression ini sendiri memiliki interpretasi yang mudah dalam bebagai kasus, salah satunya pemasaran bank. Logistic regression juga merupakan suastu metode yang baik untuk memprediksi probabilitas, memiliki performa yang baik untuk data yang telah terstruktur. Dalam pemasaran bank ini salah satu metode yang sesuai sehingga dapa tmenjadi alat yang efisien untuk analisis prediktif.

## **BAB III**

## **HASIL DAN ANALISIS**

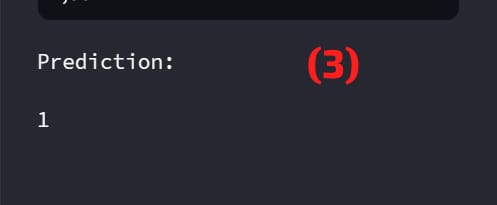
## **USER MANUAL GUIDE**



1. Dapat dilihat pada tahap pertama dengan memilih algoritma yang ingin digunakan, yaitu Random Forest, Decision Tree, dan Logistic Regression

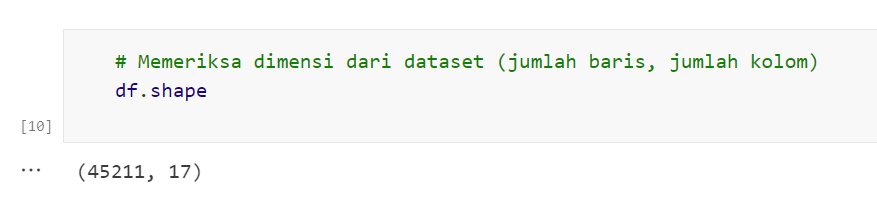


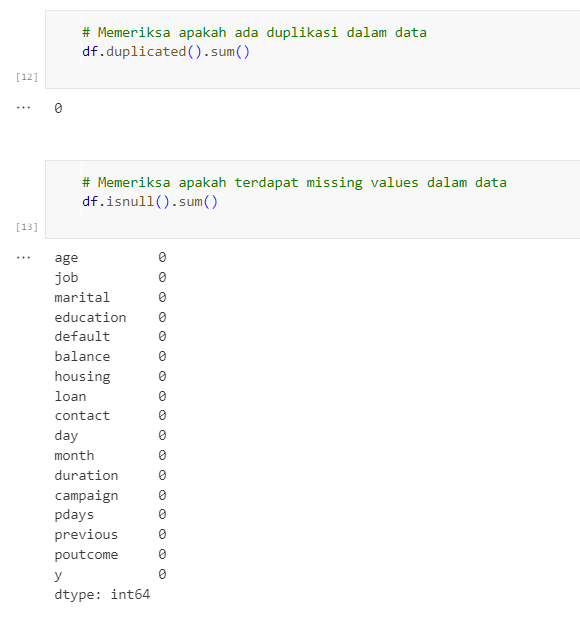
1. Pada tahap yang kedua ini dapat mengukur prediction yang ingin digunakan



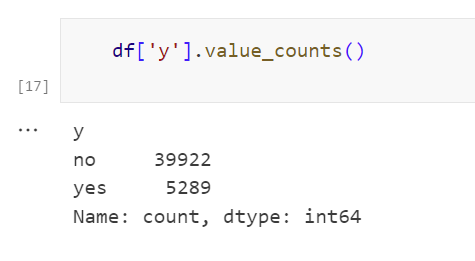
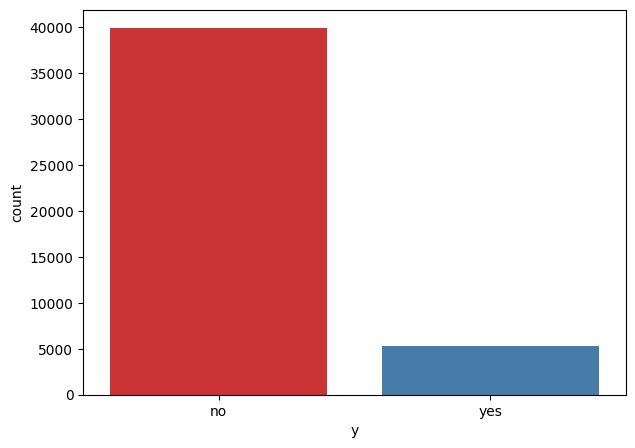
1. Tahap ketiga ini menampilkan hasil dari model dan pengukuran yang telah kita gunakan sehingga akan mengeluarkan hasil 1 untuk “yes” dan 0 untuk “no”.

## **LISTING CODE**

1. Diatas adalah library yang digunakan dalam project ini.
2. Pada gambar diatas menggunakan dataset ‘’bank-full.csv’’ diketahui 45.211 data dengan 17 kolom.



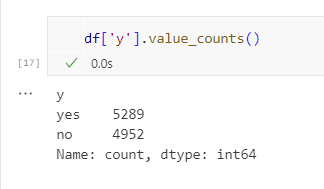
1. Diatas dilakukan data cleaning dengan memeriksa apakah terdaapt duplikasi data dan missing value.

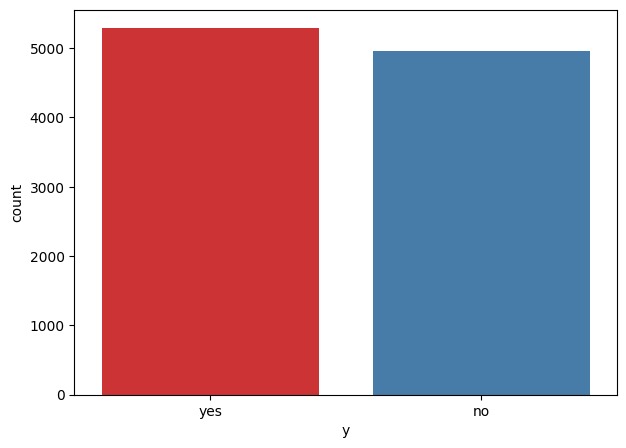


1. Dengan memeriksa frekuensi nilai pada variabel y diketahui ‘’no’’ sebanyak 39.922 dan ‘’yes’’ sebanyak 5.289 dan disebelah kanan terdapat visualisasi data yang terlihat ketimpangan antara nilai ‘’no’’ dan ‘’yes’’ tersebut.

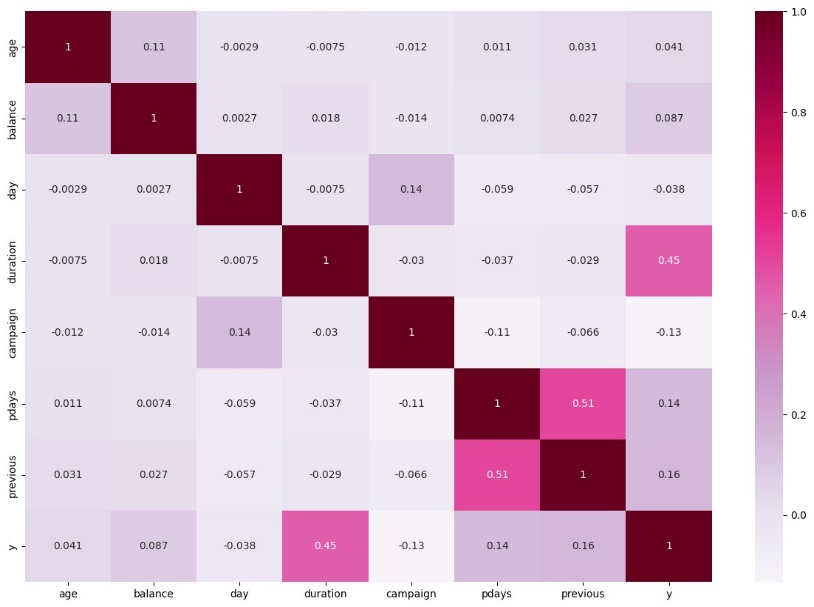


1. Pada code diatas terdapat code resample untuk membuat data tersebut dari inbalance menjadi balance, dimana ‘’no’’ sebanyak 5.300 dan ‘’yes’’ sebanyak 5.289.

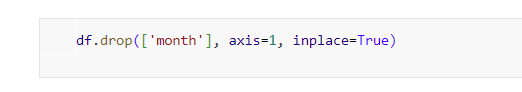


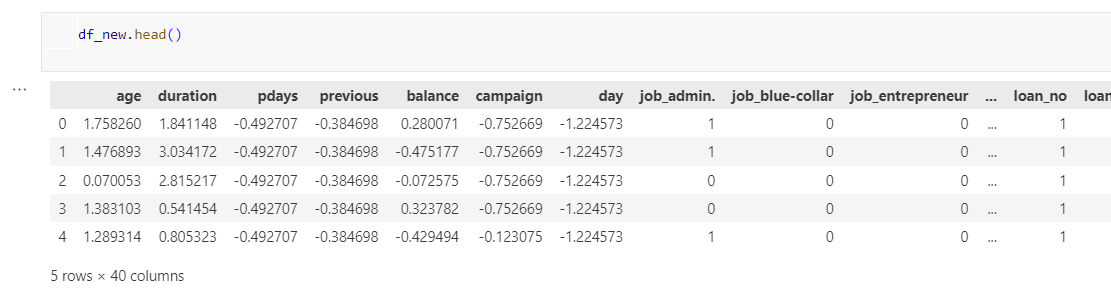


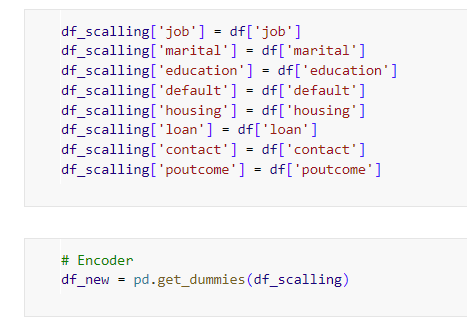
1. Ini merupakan nilai frekuensi y setelah di resample.



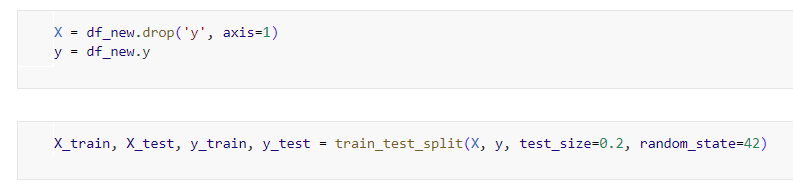
1. Dapat dilihat korelasinya terhadap variabel y menggunakan heatmap.



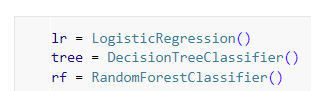
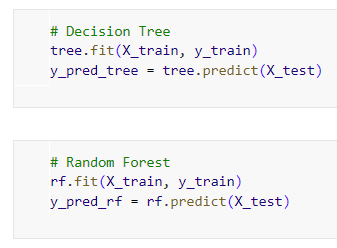
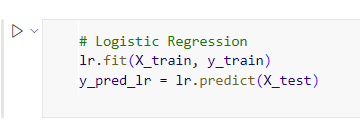
1. Kami melakukan preprocessing dnegan penghapusan kolom yang tidak berpengaruh terhadap variabel y.

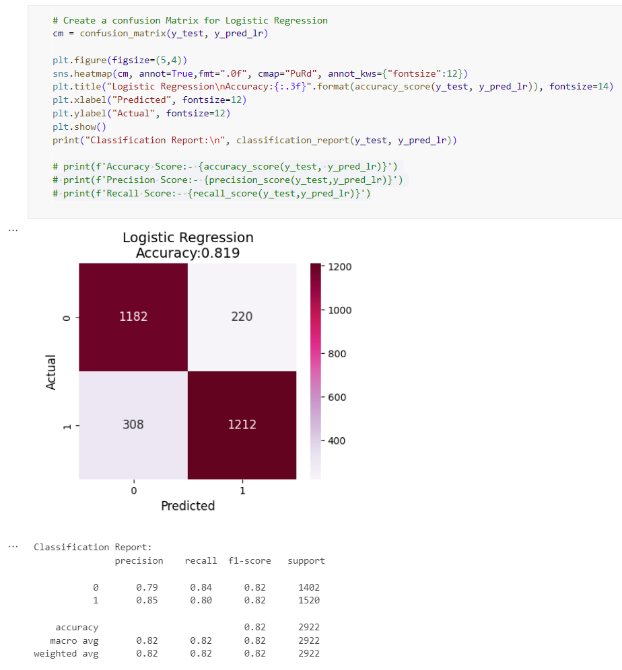


1. Kemudian data tersebut kami encoder untuk disimpan dalam format tertentu agar mudah dianalisis.



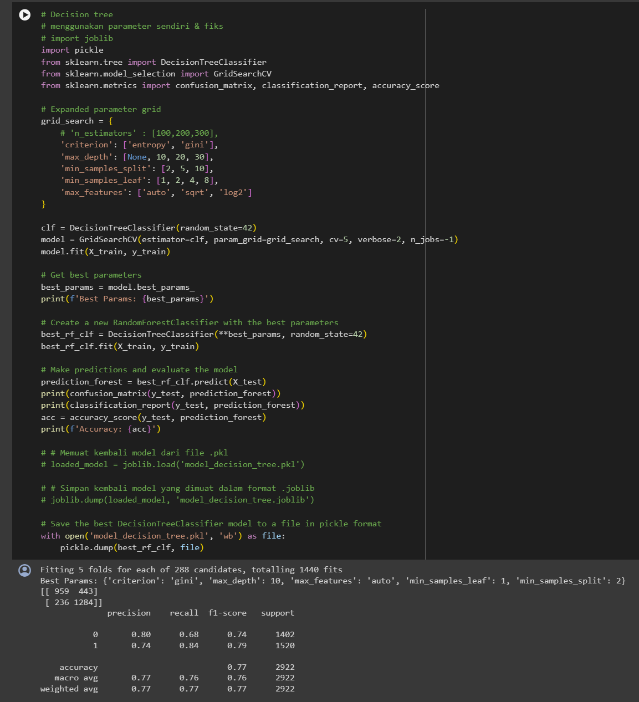
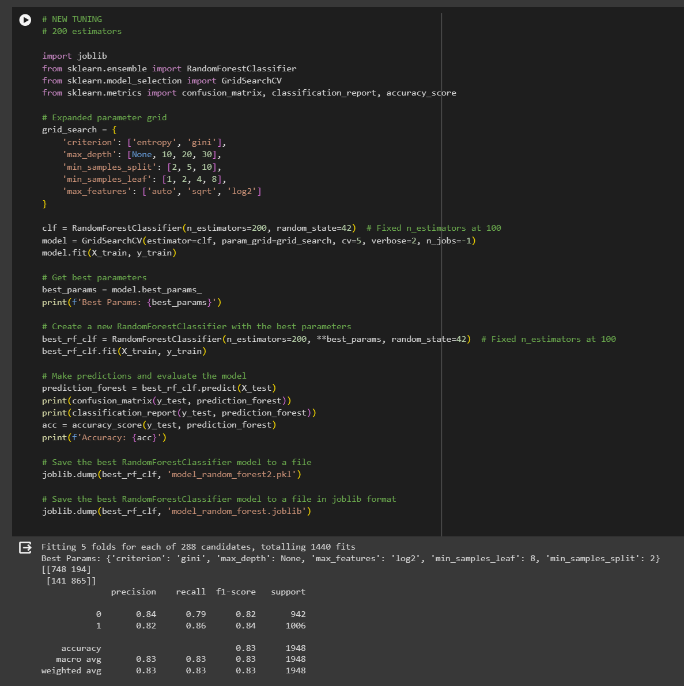
1. Data tersebut di split.

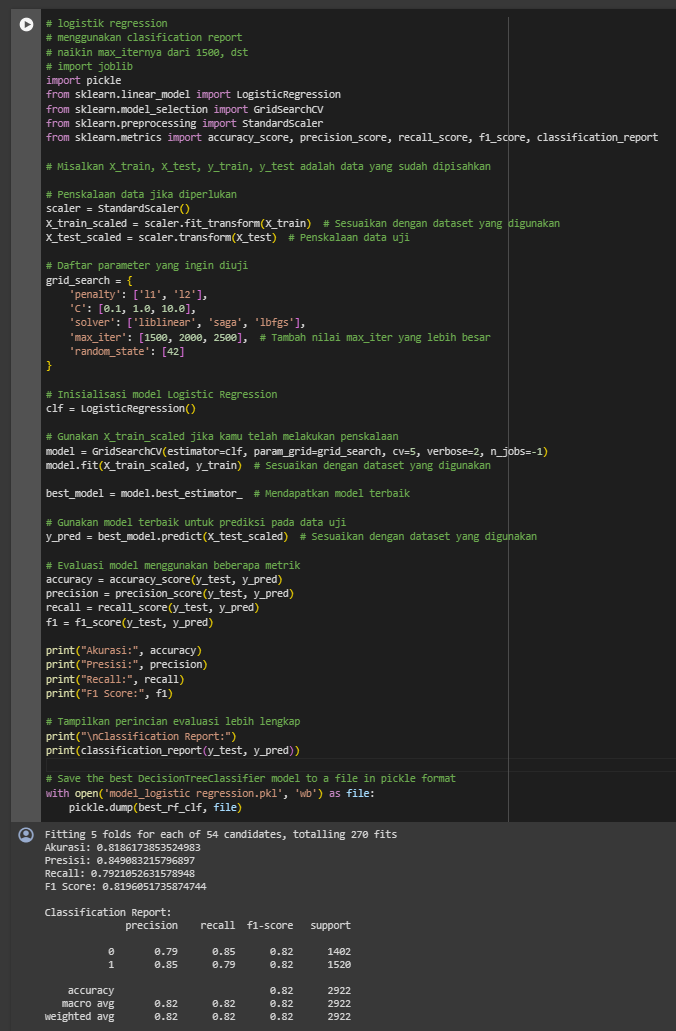


1.  Pada code diatas dilakukannya pendefinisian dan pelatihan model.



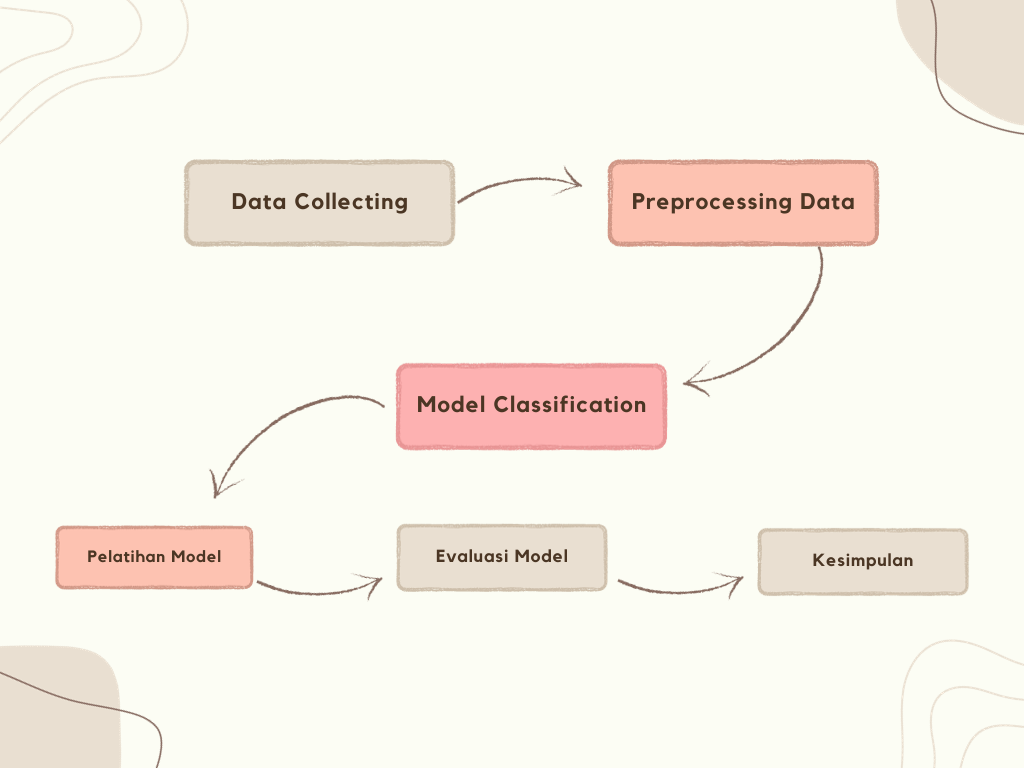
1. Diatas merupakan hasil evaluasi dari ketiga model





1. Code diatas merupakan hasil tuning dari Random Forest, Naïve Bayes, dan KNN

## **DIAGRAM ALIR**



## **DETAIL JOB PROJECT**

1. Ahmad Rafi Syaifudin (22031554030)

* Streamlit
* Cleaning Data
* PPT
* Undersampling
* Hyperparameter Tuning
* EDA

1. Analicia (22031554007)

* Modelling
* Split Data
* PPT
* Laporan
* Data Preprocessing
* Hyperparameter Tuning

1. Riva Dian Ardiansyah (22031554043)

* Streamlit
* Data Collecting
* PPT
* Confusion Matrix
* Feature Importance

## **BAB IV**

## **KESIMPULAN**

Dalam laporan project ini dengan melakukan prediksi peningkatan pemasaran bank. Dengan beberapa tahap dalam pengklasifikasi yang dilakukan, yaitu observing data, cleaning data, kemudian melakukan evaluasi model, dan hyperparameter tuning, melakukan evaluasi juga melihat confusion matrix nya hingga pada klasifikasi reportnya.

Berdasarkan hasil keseluruhan yang telah kami lakukan. Pada prediksi peningkatan pemasaran bank dengan menggunakan 3 algoritma diantaranya Random Forest, Logistic Regression, Decision Tree. Kami melihat perbedaan tiap model, sehingga menunjukan performa terbaik dari ketiga model algoritma. Dari ketiga algoritma didapatkan hasil terbaik dari Random Forest yang telah dilakukan resample dan tuning sehingga menghasilkan akurasi sebesar 83%.

## **REFERENSI**

[1] O. Apampa, “Evaluation of Classification and Ensemble Algorithms for Bank Customer Marketing Response Prediction,” *Journal of International Technology and Information Management*, vol. 25, no. 4, Jan. 2016, doi: 10.58729/1941-6679.1296.

[2] P. Appiahene, Y. M. Missah, and U. Najim, “Predicting Bank Operational Efficiency Using Machine Learning Algorithm: Comparative Study of Decision Tree, Random Forest, and Neural Networks,” *Advances in Fuzzy Systems*, vol. 2020, 2020, doi: 10.1155/2020/8581202.

[3] F. M. Basysyar, A. R. Dikananda, and D. A. Kurnia, “Prediction of Bank Customer Potential Using Creative Marketing Based on Exploratory Data Analysis and Decision Tree Algorithm,” *Ingenierie des Systemes d’Information*, vol. 27, no. 4, pp. 597–604, Aug. 2022, doi: 10.18280/isi.270409.

[4] F. Safarkhani and S. Moro, “Improving the accuracy of predicting bank depositor’s behavior using a decision tree,” *Applied Sciences (Switzerland)*, vol. 11, no. 19, Oct. 2021, doi: 10.3390/app11199016.

[5] M. C. Keshava, “Predictive Analysis on Bank Marketing Campaign using Machine Learning Algorithms,” *Int J Res Appl Sci Eng Technol*, vol. 7, no. 4, pp. 2664–2668, Apr. 2019, doi: 10.22214/ijraset.2019.4483.

[6] J. Alexandra and K. P. Sinaga, “Machine Learning Approaches for Marketing Campaign in Portuguese Banks,” *2021 3rd International Conference on Cybernetics and Intelligent System (ICORIS)*, Oct. 2021, doi: https://doi.org/10.1109/icoris52787.2021.9649623.

[7] F. M. Basysyar, A. R. Dikananda, and D. A. Kurnia, “Prediction of Bank Customer Potential Using Creative Marketing Based on Exploratory Data Analysis and Decision Tree Algorithm,” *Ingénierie des systèmes d information*, vol. 27, no. 4, pp. 597–604, Aug. 2022, doi: https://doi.org/10.18280/isi.270409.

[8] S. Moro, P. Cortez, and P. Rita, “A data-driven approach to predict the success of bank telemarketing,” *Decision Support Systems*, vol. 62, pp. 22–31, Jun. 2014, doi: https://doi.org/10.1016/j.dss.2014.03.001.

[9] L. Vongchalerm, “Analysis of predicting the success of the banking telemarketing campaigns by using machine learning techniques,” *norma.ncirl.ie*, Jan. 31, 2022. https://norma.ncirl.ie/6337/ (accessed Dec. 22, 2023).

‌