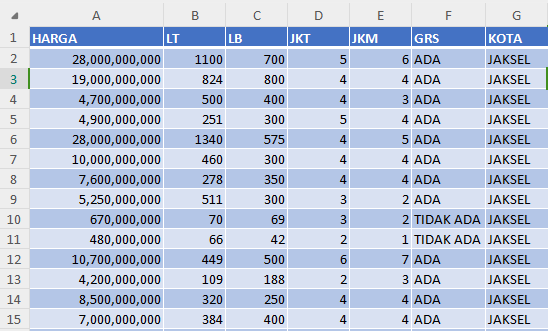
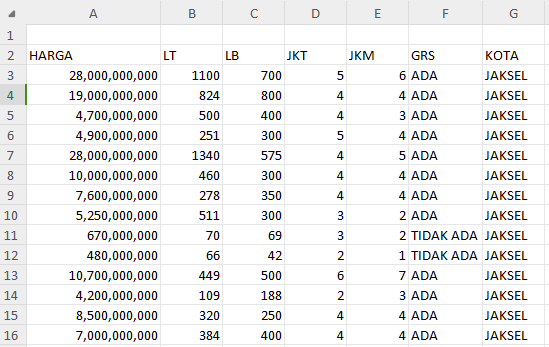
**LAPORAN UTS MACHINE LEARNING**

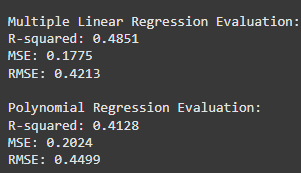
*MICHAEL PHRIGYAN HARTANTO*

*223400020*

**Tugas 1. Evaluasi Model**  
Pada tugas 1 sebelum melakukan atau menjalankan kodingan pada colab, dataset harus dirubah dahulu dari yang seperti ini, Menjadi seperti ini



Agar saat di upload ke colab tidak terjadi masalah.  
Setelah melakukan perubahan pada dataset maka akan dilakukan pembersihan data, normalisasi data, dan mengubah data kategorial menjadi numerik pada colab.

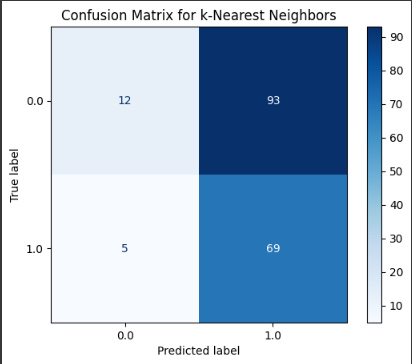
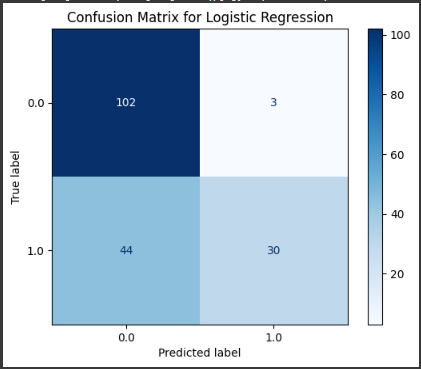
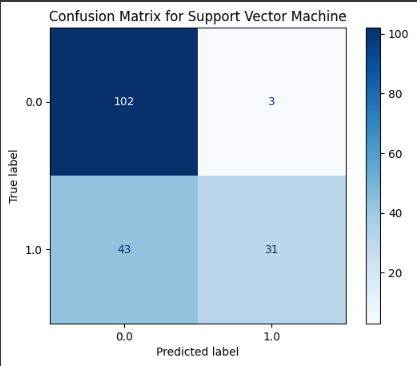
Saat sudah menjalankan colab dan juga mengupload dataset ke colab untuk di proses, hasil pemrosesan di colab menghasilkan nilai perforna masing masing model.  


Perbandingan MLR dan PR sebagai berikut  
1. Pada R-Squared dimana nilai yang lebih bagus adalah milik MLR(Multiple Linear Regression) karena lebih tinggi dari PR(Polynomial Regression) dengan selisih 700 point dimana semakin besar nilai yang dihasilkan maka lebih baik.  
2. Pada MSE dimana nilai yang lebih bagus adalah milik MLR(Multiple Linear Regression) karena lebih kecil dari PR(Polynomial Regression) dengan selisih 300 point dimana semakin kecil nilai yang dihasilkan maka lebih baik.

1. Begitu juga pada RMSE dimana nilai yang lebih bagus adalah milik MLR(Multiple Linear Regression) karena lebih kecil dari PR(Polynomial Regression) dengan selisih 200 point dimana semakin kecil nilai yang dihasilkan maka lebih baik.  
   **Tugas 2. Evaluasi Model  
   TRAIN.CSV**

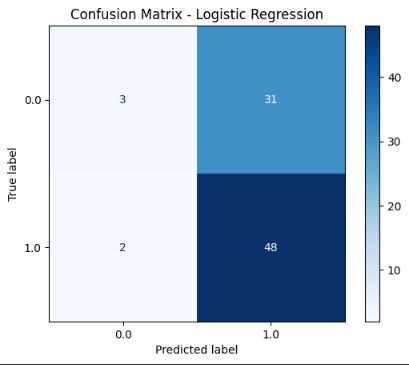
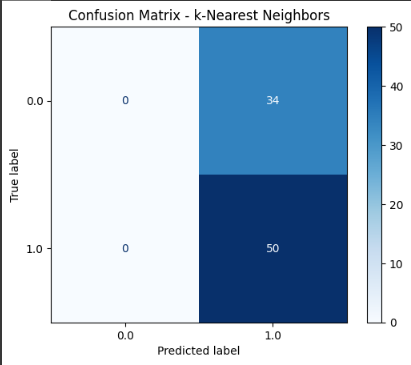
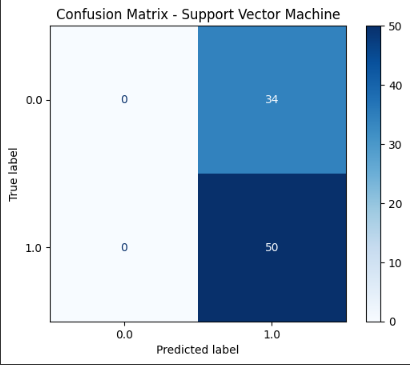
Pada dataset train.csv ini hasil nilai Akurasi, Presisi,Recall, dan F1-Score seperti pada tabel berikut

|  |  |  |  |
| --- | --- | --- | --- |
|  | Logistic Regression | KNN | SVM |
| Accuracy | 0.7374 | 0.4525 | 0.7430 |
| Precision | 0.9091 | 0.4259 | 0.9118 |
| Recall | 0.4054 | 0.9324 | 0.4189 |
| F1-Score | 0.5607 | 0.5847 | 0.5741 |

Dengan Confusion Matrix pada setiap model  
**TEST.CSV**

Pada dataset test.csv, hasil nilai Akurasi, Presisi,Recall, dan F1-Score seperti pada tabel berikut

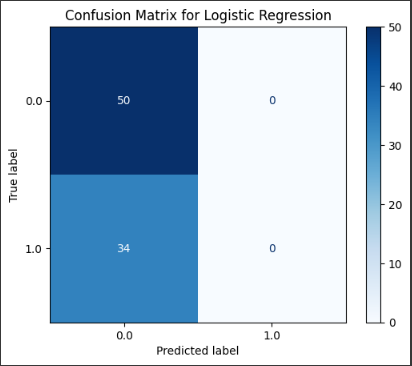
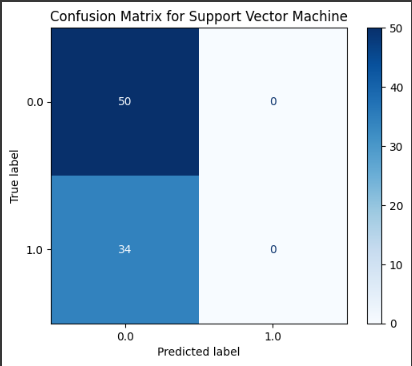
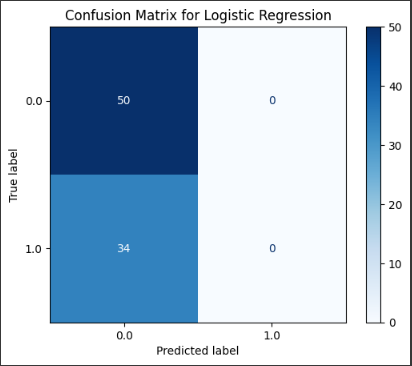
|  |  |  |  |
| --- | --- | --- | --- |
|  | Logistic Regression | KNN | SVM |
| Accuracy | 0.6071 | 0.5952 | 0.5952 |
| Precision | 0.6076 | 0.5952 | 0.5952 |
| Recall | 0.9600 | 1.000 | 1.000 |
| F1-Score | 0.7442 | 0.7463 | 0.7643 |

Dengan Confusion Matrix pada setiap model

**Gender\_Submission.csv**

Pada dataset gender\_submission.csv, hasil nilai Akurasi, Presisi,Recall, dan F1-Score seperti pada tabel berikut

|  |  |  |  |
| --- | --- | --- | --- |
|  | Logistic Regression | KNN | SVM |
| Accuracy | 0.5952 | 0.6310 | 0.5952 |
| Precision | 0.0000 | 0.6000 | 0.0000 |
| Recall | 0.0000 | 0.2647 | 0.0000 |
| F1-Score | 0.0000 | 0.3673 | 0.0000 |

Dengan Confusion Matrix pada setiap model

Sehingga untuk Tugas 1 model yang terbaik untuk dipakai adalah model Multiple Linear Regression, dan untuk Tugas 2 dengan dataset Train.csv model terbaik untuk dipakai adalah SVM, model terbaik untuk dipakai pada dataset test.csv adalah SVM, Logistic, dan KNN. Model terbaik untuk dipakai pada dataset gender\_submission adalah KNN