**Perbandingan Model Auto ML H2O dengan Berbagai Model ML pada Library Scikit Learn**

1. **Data**

Data yang digunakan merupakan kumpulan kalimat atau teks yang telah memiliki label sentimen pada masing-masing kalimat baik bernilai positif, negatif, maupun netral.

1. **Modelling**

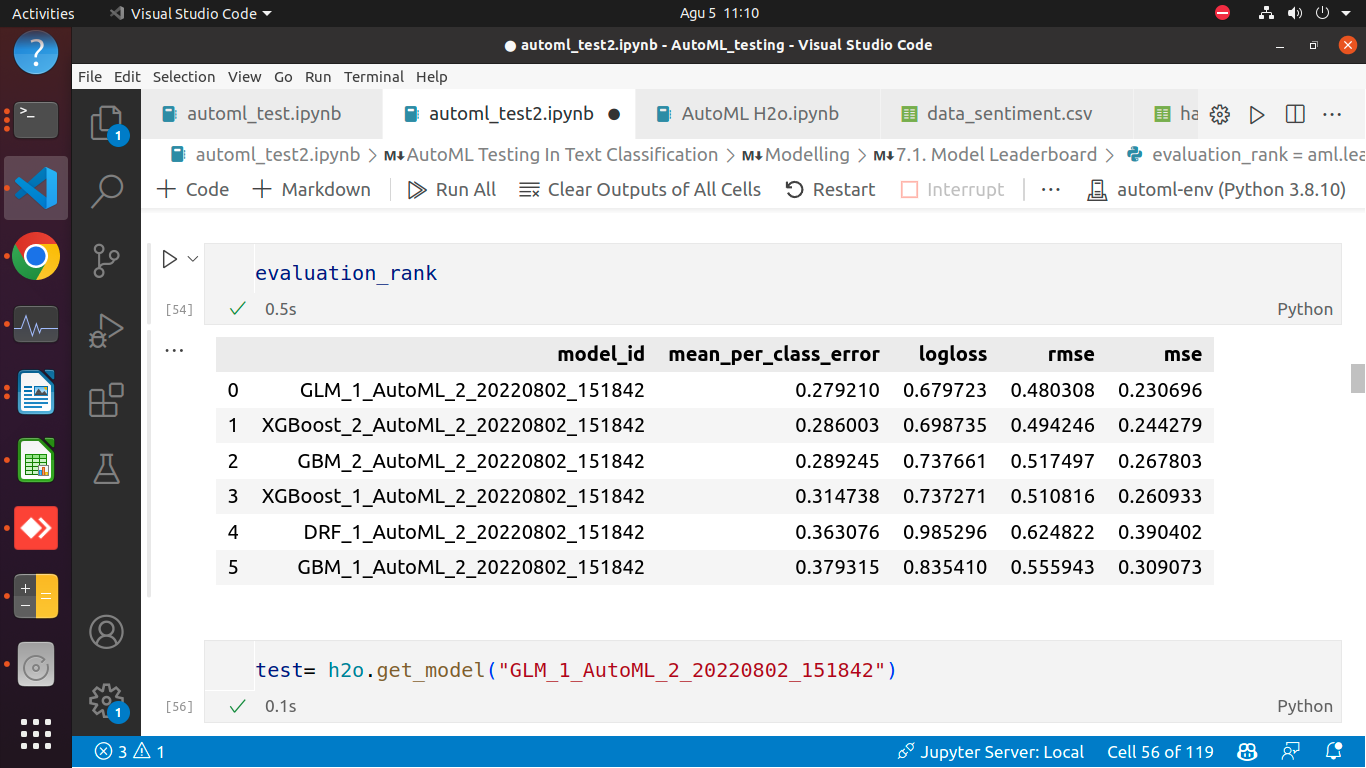
* 1. **Auto ML H2O**

Mempelajari nilai sentimen berdasarkan data-data yang diproses melalui Auto ML milik H2O. Dimana *library* ini akan menjalankan model model yang tersedia dari mereka secara otomatis.

* + 1. **Pemodelan**

Pemodelan menggunakan banyak model hingga **Maksimal** 15 model. Namun begitu, model yang ternyata berhasil dijalankan bisa saja kurang dikarenakan maksimal waktu *training* model hanya berkisar 18000 detik atau setara dengan 5 jam.

Berikut adalah model yang telah berhasil dijalankan:

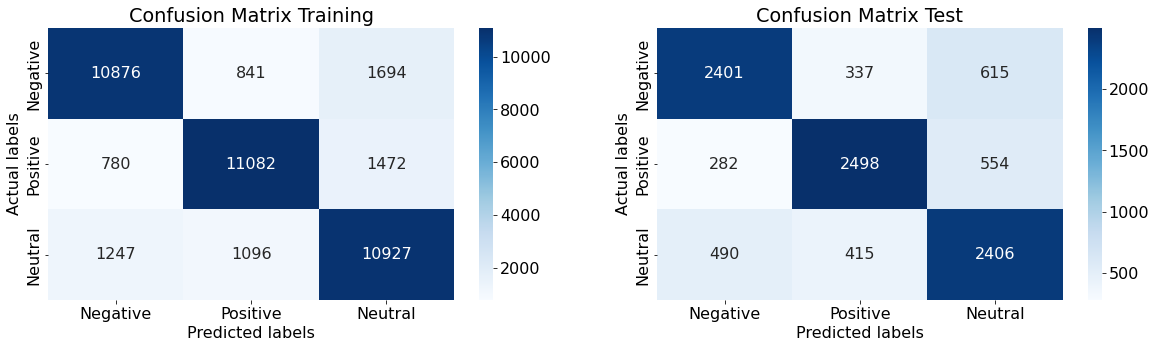


Berdasarkan tabel tersebut, ternyata model yang berhasil dijalankan hanya total 6 model saja. Program tampaknya sudah menyentuh limit waktu yang diberikan sebanyak 5 jam alih-alih menjalankan maksimal jumlah model (15). Untuk hasil evaluasi tiap model, dapat dilihat bahwa model dengan id **GLM\_1\_AutoML \_2\_20220802\_151842** merupakan model terbaik dengan memiliki nilai loss paling rendah dibandingkan model lainnya baik dilihat dari rmse, mse, logloss, ataupun mean per class.

* + 1. **Evaluasi Nilai Akurasi**

Akurasi yang didapatkan oleh model Multinomial NB untuk masing-masing kelompok data berkisar 82% untuk data *Training* dan 73% untuk data *Test*. Model sudah cukup baik untuk memprediksi data *Training* namun masih kurang dalam mengenal data *Test* dimana selisih antara evaluasi data *Training* dan *Test* cukup jauh berkisar 9%. Model masih dikatakan **Overfitting** namun kasus ini sangatlah wajar pada pengolahan data NLP, ini dikarenakan data NLP mudah mengalami **Ambiguitas**(perbedaan makna tiap kata yang bergantung dengan konteks kalimat dalam kata tersebut) dan juga terlebih kosakata dalam suatu bahasa yang sangatlah beragam. Sehingga jika model ingin mengenal data lebih baik, maka butuh penelitian dan pengenalan pola yang lebih mendalam agar model yang dihasilkan dapat mengenal pola data ataupun kata yang diberikan dari data *Training* ataupun data *Test*.

* + 1. **Confusion Matrix**

****

Berdasarkan *Confusion Matrix* di atas, kita mendapatkan informasi bahwa:

* Label **Negative Training Set**
  + True Positive(**TP**) = 10876 data
  + False Positive(**FP**) = 2027 data, (780+1247)
  + True Negative(**TN**) = 22009 data, (11082+10927)
  + False Negative(**FN**) = 2535 data, (841+1694)
* Label **Positive Training Set**
  + True Positive(**TP**) = 11082 data
  + False Positive(**FP**) = 1937 data, (841+1096)
  + True Negative(**TN**) = 21803 data, (10876+10927)
  + False Negative(**FN**) = 2252 data, (780+1472)
* Label **Neutral Training Set**
  + True Positive(**TP**) = 10927 data
  + False Positive(**FP**) = 3166 data, (1694+1472)
  + True Negative(**TN**) = 21958 data, (10876+11082)
  + False Negative(**FN**) = 2343 data, (1247+1096)
* Label **Negative Test Set**
  + True Positive(**TP**) = 2401 data
  + False Positive(**FP**) = 772 data, (282+490)
  + True Negative(**TN**) = 4904 data, (2498+2406)
  + False Negative(**FN**) = 952 data, (337+615)
* Label **Positive Test Set**
  + True Positive(**TP**) = 2498 data
  + False Positive(**FP**) = 752 data, (337+415)
  + True Negative(**TN**) = 4807 data, (2401+2406)
  + False Negative(**FN**) = 2252 data, (282+554)
* Label **Neutral Test Set**
  + True Positive(**TP**) = 2406 data
  + False Positive(**FP**) = 1169 data, (615+554)
  + True Negative(**TN**) = 4899 data, (10876+11082)
  + False Negative(**FN**) = 905 data, (490+415)
  1. **Multinomial NB**
     1. **Pemodelan**

Setelah melewati tahap *Hyperparameter Tuning*, model Multinomial NB yang terbaik dijalankan dengan menggunakan parameter nilai alpha = 0.3.

* + 1. **Evaluasi Nilai Akurasi**

Akurasi yang didapatkan oleh model Multinomial NB untuk masing-masing kelompok data berkisar 74% untuk data *Training* dan 70% untuk data *Test*. Meskipun selisih persentase tidak besar hanya berkisar 4%, namun model belum memenuhi kualifikasi **Good Fit** dimana pada umumnya dibutuhkan minimum nilai akurasi sebesar 80%.

* + 1. **Confusion Matrix**

****

Berdasarkan *Confusion Matrix* di atas, kita mendapatkan informasi bahwa:

* Label **Negative Training Set**
  + True Positive(**TP**) = 10458 data
  + False Positive(**FP**) = 4332 data
  + True Negative(**TN**) = 19036 data
  + False Negative(**FN**) = 2953 data
* Label **Positive Training Set**
  + True Positive(**TP**) = 10090 data
  + False Positive(**FP**) = 2798 data
  + True Negative(**TN**) = 19404 data
  + False Negative(**FN**) = 3244 data
* Label **Neutral Training Set**
  + True Positive(**TP**) = 8946 data
  + False Positive(**FP**) = 3391 data
  + True Negative(**TN**) = 20548 data
  + False Negative(**FN**) = 4324 data
* Label **Negative Test Set**
  + True Positive(**TP**) = 2472 data
  + False Positive(**FP**) = 1193 data
  + True Negative(**TN**) = 4489 data
  + False Negative(**FN**) = 881 data
* Label **Positive Test Set**
  + True Positive(**TP**) = 2405 data
  + False Positive(**FP**) = 844 data
  + True Negative(**TN**) = 4556 data
  + False Negative(**FN**) = 929 data
* Label **Neutral Test Set**
  + True Positive(**TP**) = 2084 data
  + False Positive(**FP**) = 1227 data
  + True Negative(**TN**) = 4877 data
  + False Negative(**FN**) = 1000 data
  1. **Logistic Regression**
     1. **Pemodelan**

Setelah melewati tahap *Hyperparameter Tuning*, model Logistic Regression yang terbaik dijalankan dengan menggunakan parameter C = 1, class\_weight = None, penalty = l2, solver = lbfgs.

* + 1. **Evaluasi Nilai Akurasi**

Akurasi yang didapatkan oleh model *Logistic Regression* untuk masing-masing kelompok data berkisar 79% untuk data *Training* dan 73% untuk data *Test*. Akurasi pada training mendekati syarat minimum(80%) namun sedikit **Overfitting**, dimana selisih Akurasi pada *Training* maupun *Testing* berkisar 6% dari maksimal selisih yang ditoleransi sebesar 5%.

* + 1. **Confusion Matrix**

****

Berdasarkan *Confusion Matrix* di atas, kita mendapatkan informasi bahwa:

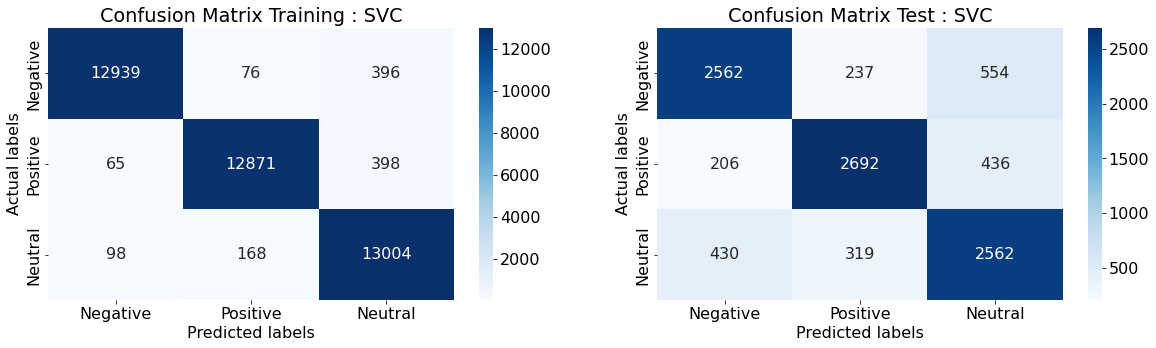
* Label **Negative Training Set**
  + True Positive(**TP**) = 10523 data
  + False Positive(**FP**) = 2472 data
  + True Negative(**TN**) = 21244 data
  + False Negative(**FN**) = 2888 data
* Label **Positive Training Set**
  + True Positive(**TP**) = 10090 data
  + False Positive(**FP**) = 2180 data
  + True Negative(**TN**) = 21100 data
  + False Negative(**FN**) = 2667 data
* Label **Neutral Training Set**
  + True Positive(**TP**) = 10577 data
  + False Positive(**FP**) = 3596 data
  + True Negative(**TN**) = 21190 data
  + False Negative(**FN**) = 2693 data
* Label **Negative Test Set**
  + True Positive(**TP**) = 2423 data
  + False Positive(**FP**) = 779 data
  + True Negative(**TN**) = 4893 data
  + False Negative(**FN**) = 930 data
* Label **Positive Test Set**
  + True Positive(**TP**) = 2481 data
  + False Positive(**FP**) = 722 data
  + True Negative(**TN**) = 4835 data
  + False Negative(**FN**) = 853 data
* Label **Neutral Test Set**
  + True Positive(**TP**) = 2412 data
  + False Positive(**FP**) = 1181 data
  + True Negative(**TN**) = 4904 data
  + False Negative(**FN**) = 899 data
  1. **SVC**
     1. **Pemodelan**

Setelah melewati tahap *Hyperparameter Tuning*, model SVC yang terbaik dijalankan dengan menggunakan parameter C = 10, kernel = rbf.

* + 1. **Evaluasi Nilai Akurasi**

Akurasi yang didapatkan oleh model SVC untuk masing-masing kelompok data berkisar 97% untuk data *Training* dan 78% untuk data *Test*.

* + 1. **Confusion Matrix**

****

Berdasarkan *Confusion Matrix* di atas, kita mendapatkan informasi bahwa:

* Label **Negative Training Set**
  + True Positive(**TP**) = 12939 data
  + False Positive(**FP**) = 163 data
  + True Negative(**TN**) = 25875 data
  + False Negative(**FN**) = 472 data
* Label **Positive Training Set**
  + True Positive(**TP**) = 12871 data
  + False Positive(**FP**) = 244 data
  + True Negative(**TN**) = 25943 data
  + False Negative(**FN**) = 463 data
* Label **Neutral Training Set**
  + True Positive(**TP**) = 13004 data
  + False Positive(**FP**) = 794 data
  + True Negative(**TN**) = 25810 data
  + False Negative(**FN**) = 266 data
* Label **Negative Test Set**
  + True Positive(**TP**) = 2562 data
  + False Positive(**FP**) = 636 data
  + True Negative(**TN**) = 5254 data
  + False Negative(**FN**) = 791 data
* Label **Positive Test Set**
  + True Positive(**TP**) = 2692 data
  + False Positive(**FP**) = 556 data
  + True Negative(**TN**) = 5124 data
  + False Negative(**FN**) = 642 data
* Label **Neutral Test Set**
  + True Positive(**TP**) = 2562 data
  + False Positive(**FP**) = 990 data
  + True Negative(**TN**) = 5254 data
  + False Negative(**FN**) = 749 data
  1. **XGBoost Classifier**
     1. **Pemodelan**

Setelah melewati tahap *Hyperparameter Tuning*, model XGBoost Classifier yang terbaik dijalankan dengan menggunakan parameter colsample\_bytree = 0.7, gamma = 0, max\_depth = 4, n\_estimators = 150, subsample = 0.7.

* + 1. **Evaluasi Nilai Akurasi**

Akurasi yang didapatkan oleh model XGBoost Classifier untuk masing-masing kelompok data berkisar 73% untuk data *Training* dan 68% untuk data *Test*. Performa yang didapatkan model ini paling rendah dibandingkan model-model lainnya. Kemungkinan model masih membutuhkan *Tuning* yang lebih beragam untuk mendapatkan hasil yang jauh lebih baik.

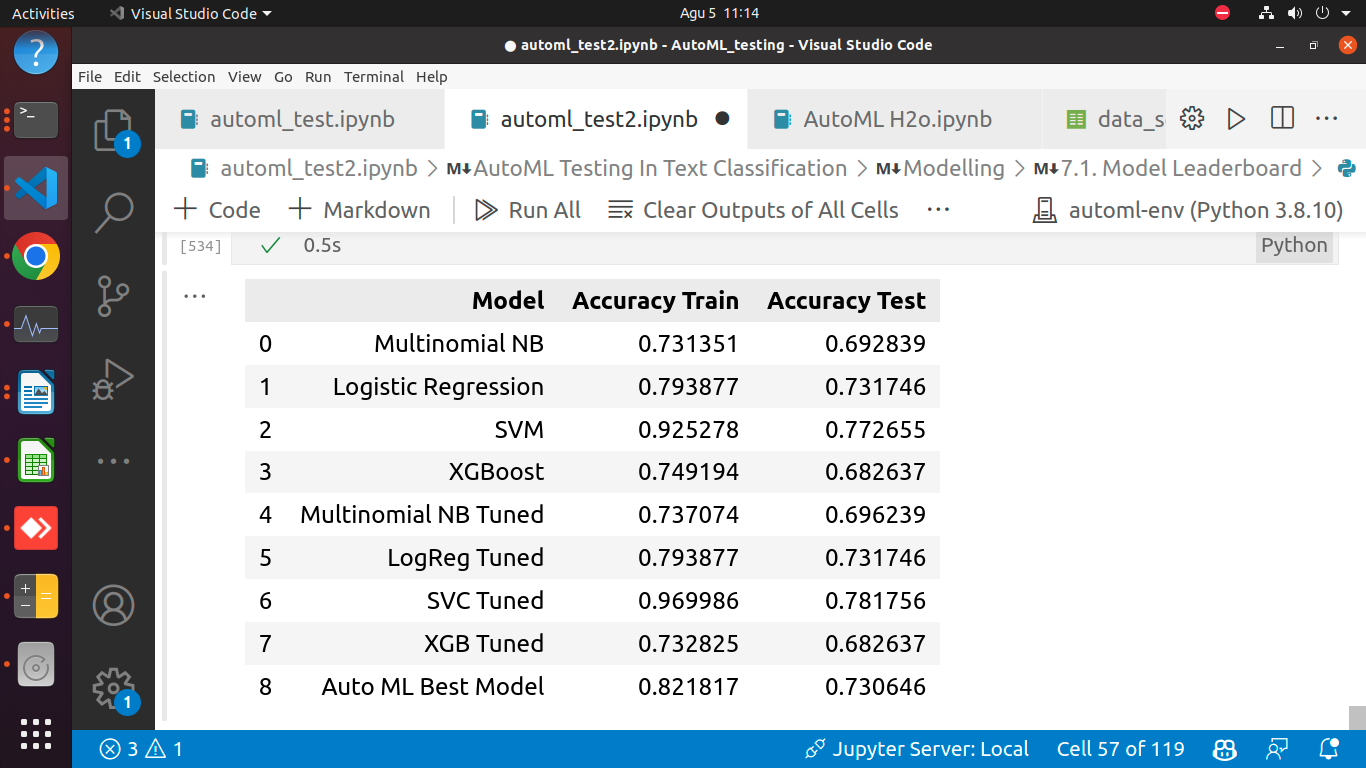
* + 1. **Confusion Matrix**

****

Berdasarkan *Confusion Matrix* di atas, kita mendapatkan informasi bahwa:

* Label **Negative Training Set**
  + True Positive(**TP**) = 9053 data
  + False Positive(**FP**) = 2266 data
  + True Negative(**TN**) = 20271 data
  + False Negative(**FN**) = 4358 data
* Label **Positive Training Set**
  + True Positive(**TP**) = 9237 data
  + False Positive(**FP**) = 2146 data
  + True Negative(**TN**) = 20087 data
  + False Negative(**FN**) = 4097 data
* Label **Neutral Training Set**
  + True Positive(**TP**) = 11034 data
  + False Positive(**FP**) = 6279 data
  + True Negative(**TN**) = 18290 data
  + False Negative(**FN**) = 2236 data
* Label **Negative Test Set**
  + True Positive(**TP**) = 2089 data
  + False Positive(**FP**) = 723 data
  + True Negative(**TN**) = 4736 data
  + False Negative(**FN**) = 1264 data
* Label **Positive Test Set**
  + True Positive(**TP**) = 2163 data
  + False Positive(**FP**) = 723 data
  + True Negative(**TN**) = 4662 data
  + False Negative(**FN**) = 1264 data
* Label **Neutral Test Set**
  + True Positive(**TP**) = 2573 data
  + False Positive(**FP**) = 669 data
  + True Negative(**TN**) = 4252 data
  + False Negative(**FN**) = 1171 data

1. **Metrics Evaluation**

****

Berdasarkan tabel di atas, secara *overall* model yang paling aman dijalankan sementara ialah model **SVC** yang telah di-*tuning* dengan hasil nilai akurasi 97% untuk data *Training,* dan 78% untuk data *Test*. Meskipun selisih akurasi sangatlah jauh sehingga model masih dikategorikan **Overfit**, namun akurasi untuk data *Test* dari model paling terbaik di antara model-model lainnya termasuk model dari Auto ML milik H2O. Jika ditelusuri lebih jauh, keseluruhan model ML cenderung lebih susah untuk mengenali label **Neutral** dimana pada label ini nilai **F1-Score** lebih rendah jika dibandingkan dengan label lainnya. Dan lebih jauh lagi, keseluruhan model kecuali **Multinomial NB** rata-rata memiliki nilai precision yang paling buruk pada label **Neutral**. Hal ini berarti model mudah untuk mengeneralisasikan suatu label bernada **Neutral** dimana seharusnya label pada kata atau kalimat tersebut bernada **Negatif** ataupun **Positif.**