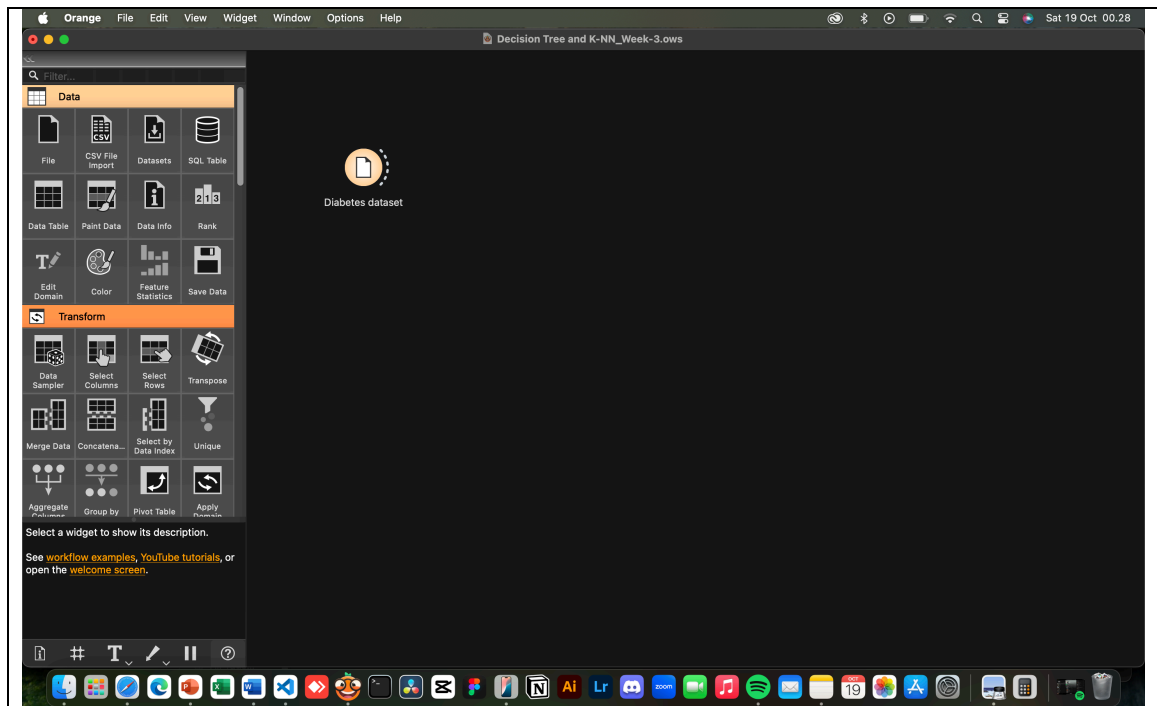


Nama	Muhammad Abyan Ridhan Siregar
NIM	1103210053
Kelas	TK-45-01

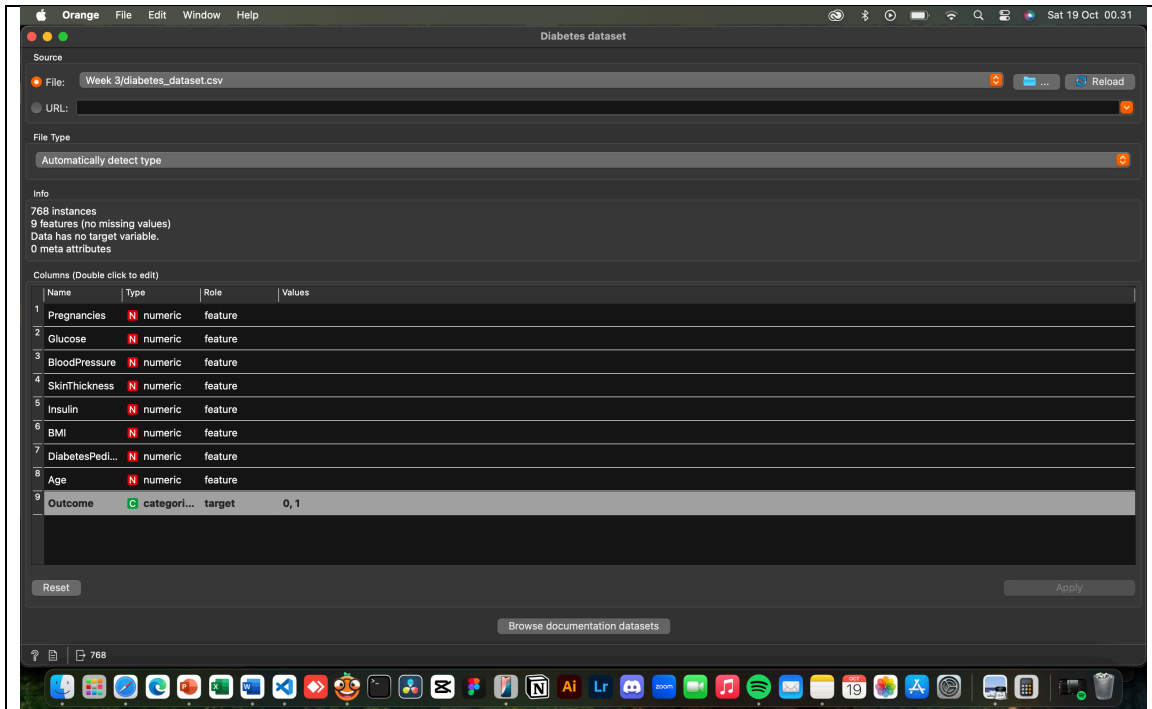
Decision Tree and K-NN using Orange Data Mining

- Step 1



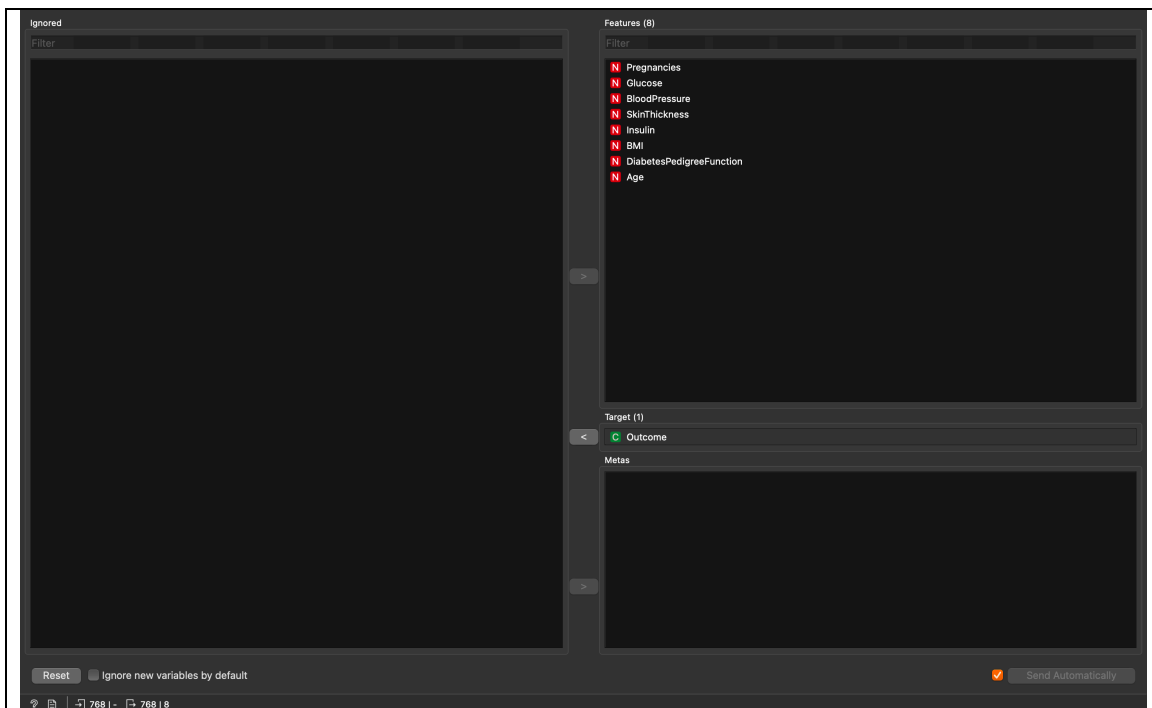
Buka orange data mining, drag n drop file dan namakan “Diabetes Dataset”

- Step 2



Kemudian import dataset yang ingin kita gunakan dengan format Nama, Type, and Role sesuai dengan yang ada di gambar di atas.

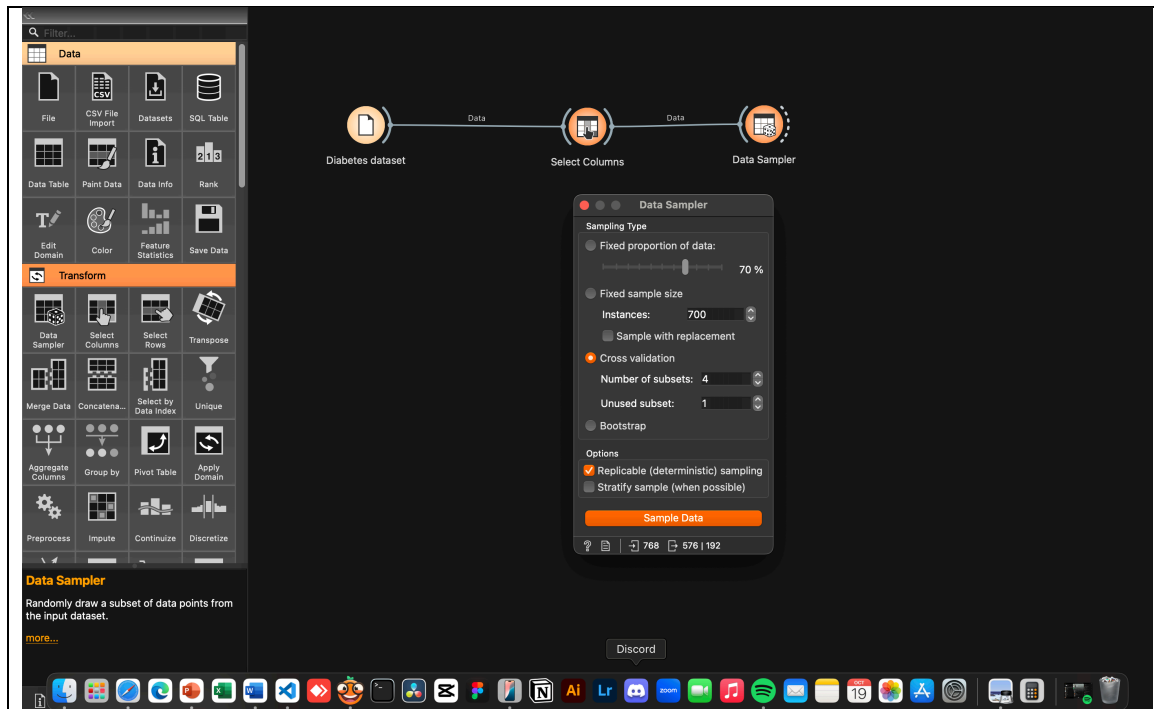
- Step 3



Seret widget **Select Columns** ke kanvas dan hubungkan dengan widget **File**. Klik dua kali widget **Select Columns**. Pastikan variabel target

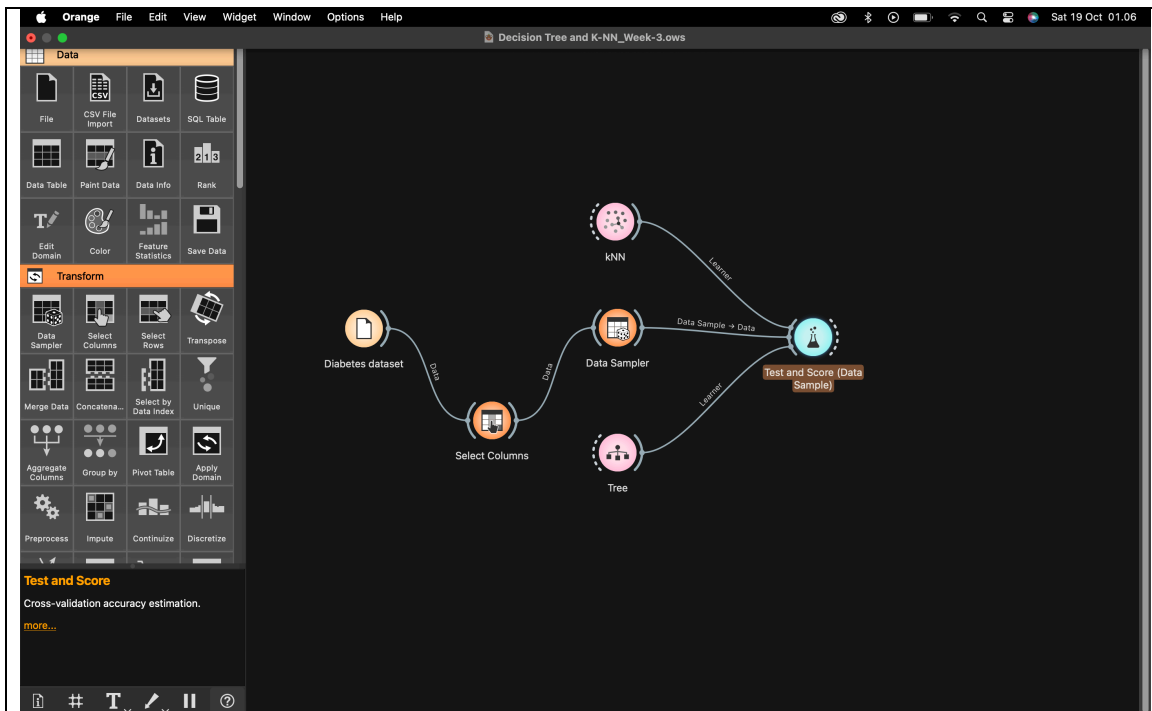
(misalnya, Outcome atau yang setara) dipilih sebagai target. Biarkan kolom lain sebagai fitur. Klik **Apply**.

- Step 4



Seret widget **Data Sampler** dan hubungkan dengan widget **Select Columns**. Klik dua kali widget **Data Sampler**. Atur ukuran sampel menjadi **70%** untuk pelatihan dan **30%** untuk pengujian. Pastikan **Replicable** dicentang agar hasilnya konsisten. Klik **Apply**.

- Step 5



Cross validation

Number of folds: 5

☒ Stratified

☐ Cross validation by feature

☐ Random sampling

Repeat train/test: 10

Training set size: 66 %

☒ Stratified

☐ Leave one out

☐ Test on train data

☐ Test on test data

Evaluation results for target 1

Model	AUC	CA	F1	Prec	Recall	MCC
Tree	0.755	0.738	0.488	0.750	0.362	0.380
kNN	0.769	0.753	0.612	0.671	0.563	0.437

Compare models by: Area under ROC curve

Negligible diff.: 0.1

	Tree	kNN
Tree		0.340
kNN	0.660	

Table shows probabilities that the score for the model in the row is higher than that of the model in the column. Small numbers show the probability that the difference is negligible.

Penjelasan Metrik Evaluasi

1. AUC (Area Under the Curve):

- o Untuk Tree: 0.755
- o Untuk kNN: 0.769

- Ini menunjukkan kemampuan model dalam membedakan antara kelas positif dan negatif. Kedua nilai menunjukkan bahwa model lebih baik dari tebakan acak, tetapi nilai 0.769 dari kNN sedikit lebih baik daripada Tree.

2. CA (Classification Accuracy):

- Untuk Tree: 0.738 (73.8%)
- Untuk kNN: 0.753 (75.3%)
- Akurasi untuk kedua model masih di bawah target 85%. kNN sedikit lebih baik daripada Tree, tetapi tetap belum memenuhi ketentuan awal.

3. F1 Score:

- Untuk Tree: 0.488
- Untuk kNN: 0.612
- kNN memiliki F1 Score yang lebih tinggi, yang menunjukkan keseimbangan yang lebih baik antara precision dan recall dibandingkan Tree.

4. Prec (Precision):

- Untuk Tree: 0.750
- Untuk kNN: 0.671
- Precision pada Tree lebih tinggi daripada kNN, menunjukkan bahwa model Tree lebih baik dalam memprediksi benar dari yang diprediksi positif. Namun, ini mungkin disebabkan oleh rendahnya recall.

5. Recall:

- Untuk Tree: 0.362
- Untuk kNN: 0.563
- kNN memiliki nilai recall yang lebih tinggi daripada Tree, yang berarti kNN lebih baik dalam menemukan semua contoh positif dalam data. Namun, keduanya masih memerlukan peningkatan agar mendekati atau melebihi nilai yang ideal.

6. MCC (Matthews Correlation Coefficient):

- Untuk Tree: 0.380
- Untuk kNN: 0.437
- kNN menunjukkan korelasi yang sedikit lebih baik dalam prediksinya dibandingkan Tree, tetapi masih jauh dari sempurna (nilai mendekati 1).

Analisis Hasil

Kesimpulan: Akurasi (CA) untuk kNN adalah 75.3%, dan Tree adalah 73.8%. Keduanya masih berada di bawah target 85%, sehingga model saat ini belum memenuhi ketentuan awal. Model kNN menunjukkan performa yang sedikit lebih baik secara keseluruhan dibandingkan Tree, tetapi masih perlu peningkatan lebih lanjut untuk mencapai target akurasi

