



# MACHINE LEARNING : MODELING USING ORANGE

---

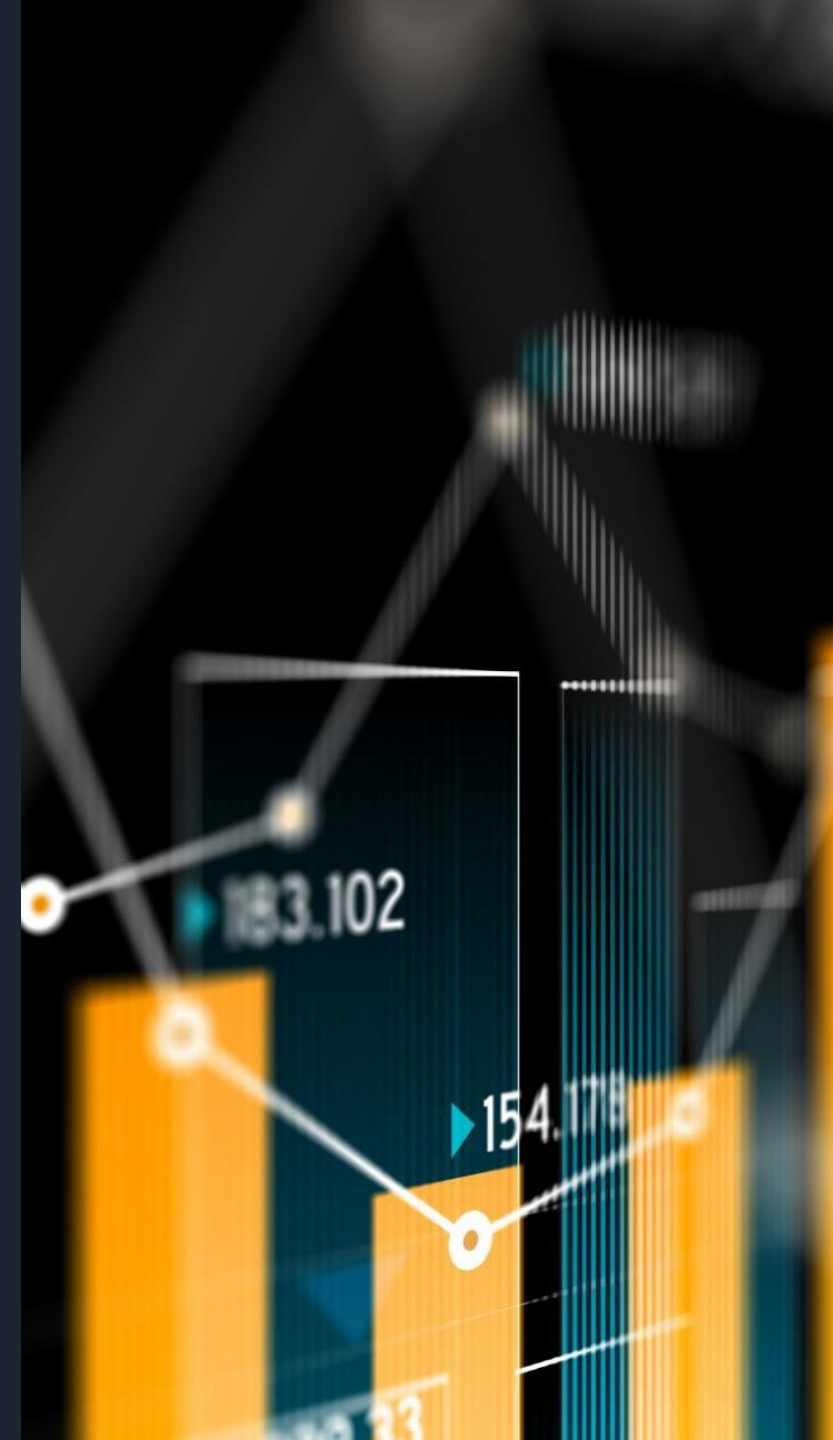
*Alvandi Damansyah*

1103192191

# ORANGE DATA MINING

---

Orange Data Mining merupakan sebuah aplikasi open-source yang menyediakan visualisasi data, machine learning, dan data analisis secara interaktif.



# LINEAR MODEL

---

Merupakan model yang meneliti antar variabel, ataupun memprediksi nilai data yang tidak diketahui dengan menggunakan nilai data lain yang terkait dan diketahui.



# DECISION TREE MODEL

---

Merupakan sebuah model yang menggunakan aturan untuk membuat keputusan dengan struktur seperti pohon yang memodelkan kemungkinan hasil, konsekuensi, ataupun resiko.

# RANDOM FOREST MODEL

---

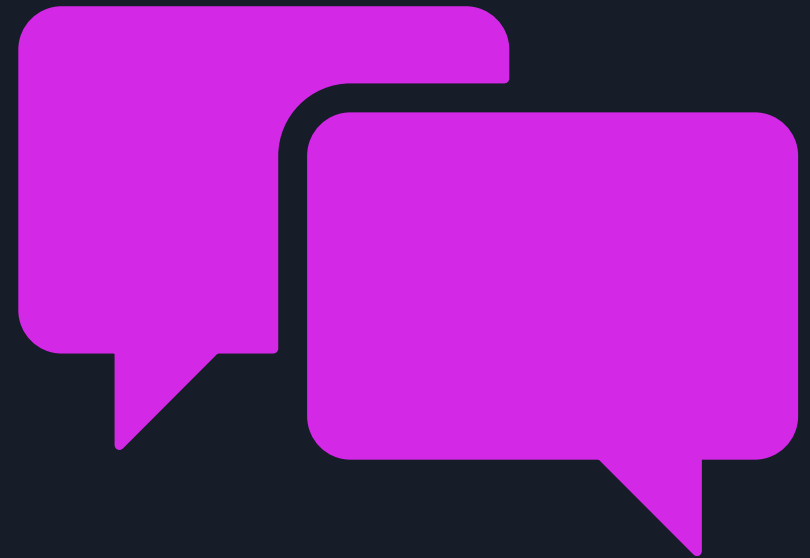
Merupakan model yang terbentuk dari beberapa decision tree yang digabungkan untuk mendapatkan prediksi yang lebih akurat.



# CHAT GPT

---

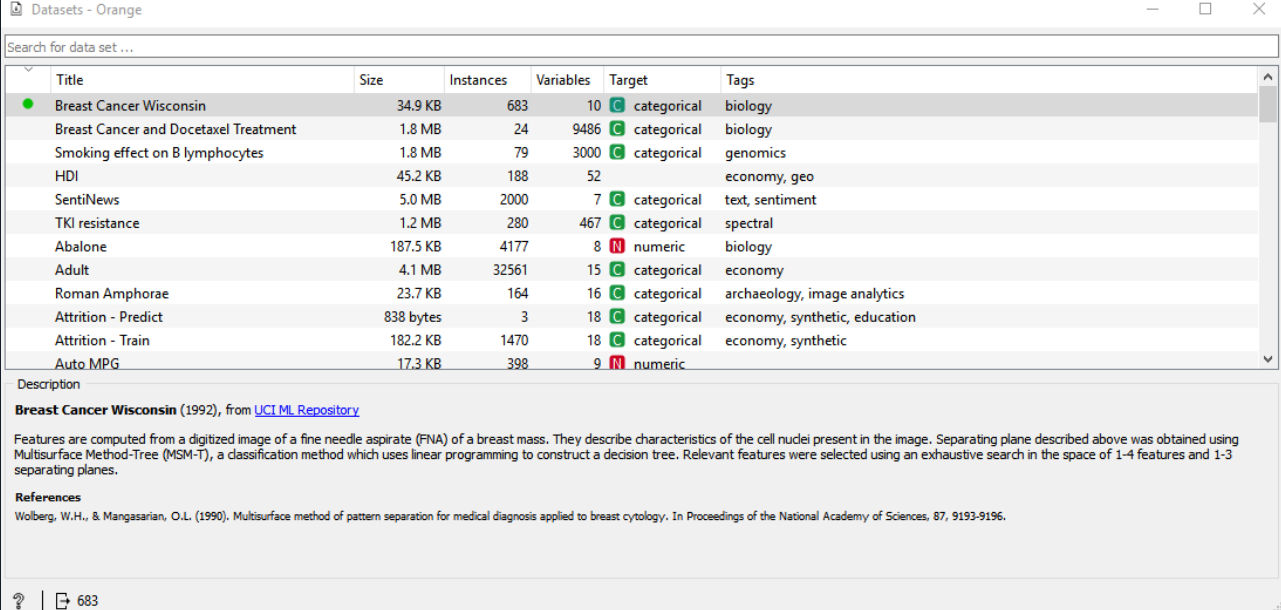
Merupakan kecerdasan buatan berbentuk chatbot yang akan menjawab pertanyaan yang diberikan.





# RESULT

Dataset yang digunakan merupakan  
*breast cancer wisconsin*



Title	Size	Instances	Variables	Target	Tags
Breast Cancer Wisconsin	34.9 KB	683	10	categorical	biology
Breast Cancer and Docetaxel Treatment	1.8 MB	24	9486	categorical	biology
Smoking effect on B lymphocytes	1.8 MB	79	3000	categorical	genomics
HDI	45.2 KB	188	52		economy, geo
SentiNews	5.0 MB	2000	7	categorical	text, sentiment
TKI resistance	1.2 MB	280	467	categorical	spectral
Abalone	187.5 KB	4177	8	numeric	biology
Adult	4.1 MB	32561	15	categorical	economy
Roman Amphorae	23.7 KB	164	16	categorical	archaeology, image analytics
Attrition - Predict	838 bytes	3	18	categorical	economy, synthetic, education
Attrition - Train	182.2 KB	1470	18	categorical	economy, synthetic
Auto MPG	17.3 KB	398	9	numeric	

Description

**Breast Cancer Wisconsin** (1992), from [UCI ML Repository](#)

Features are computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. They describe characteristics of the cell nuclei present in the image. Separating plane described above was obtained using Multisurface Method-Tree (MSM-T), a classification method which uses linear programming to construct a decision tree. Relevant features were selected using an exhaustive search in the space of 1-4 features and 1-3 separating planes.

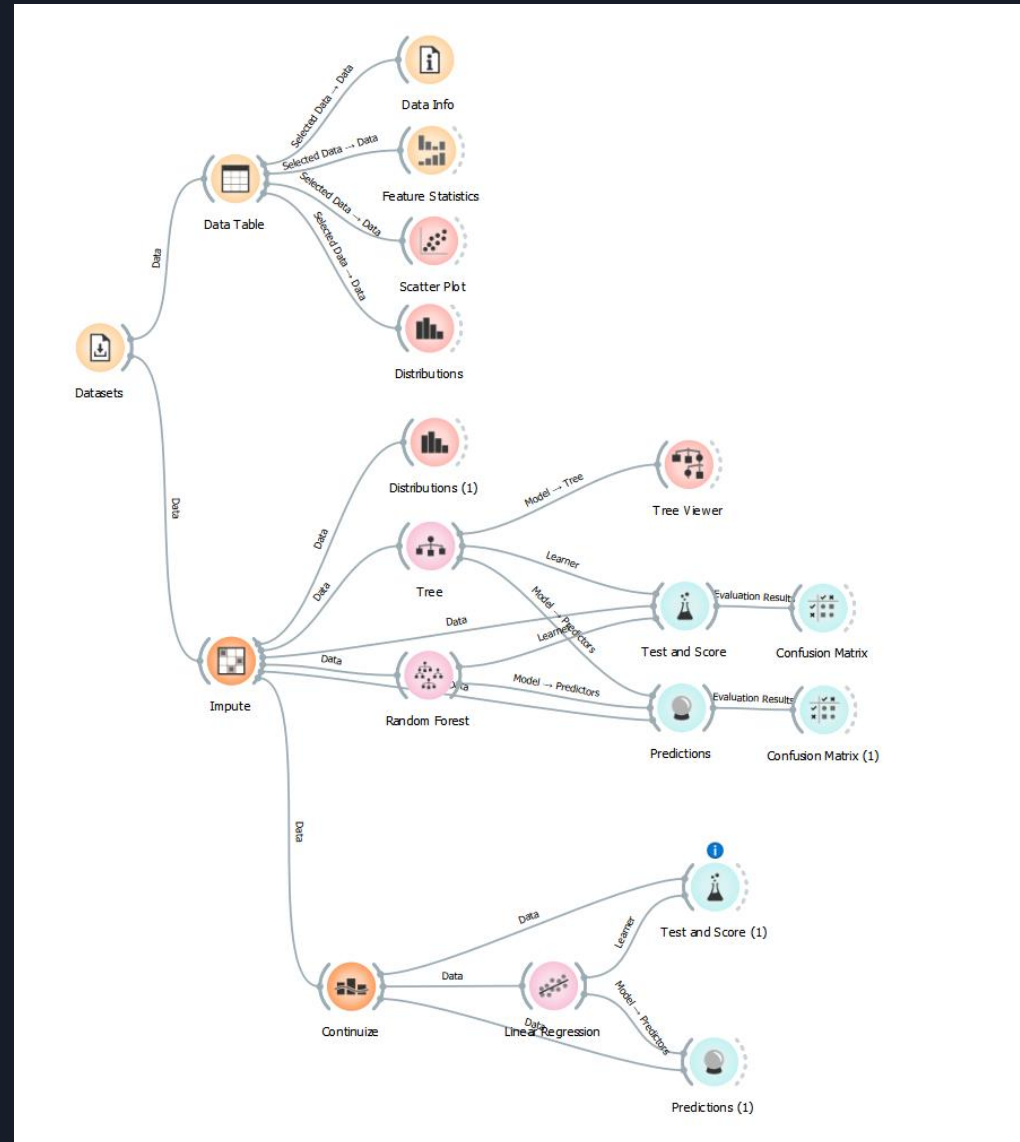
**References**

Wolberg, W.H., & Mangasarian, O.L. (1990). Multisurface method of pattern separation for medical diagnosis applied to breast cytology. In Proceedings of the National Academy of Sciences, 87, 9193-9196.

? | 683

# RESULT

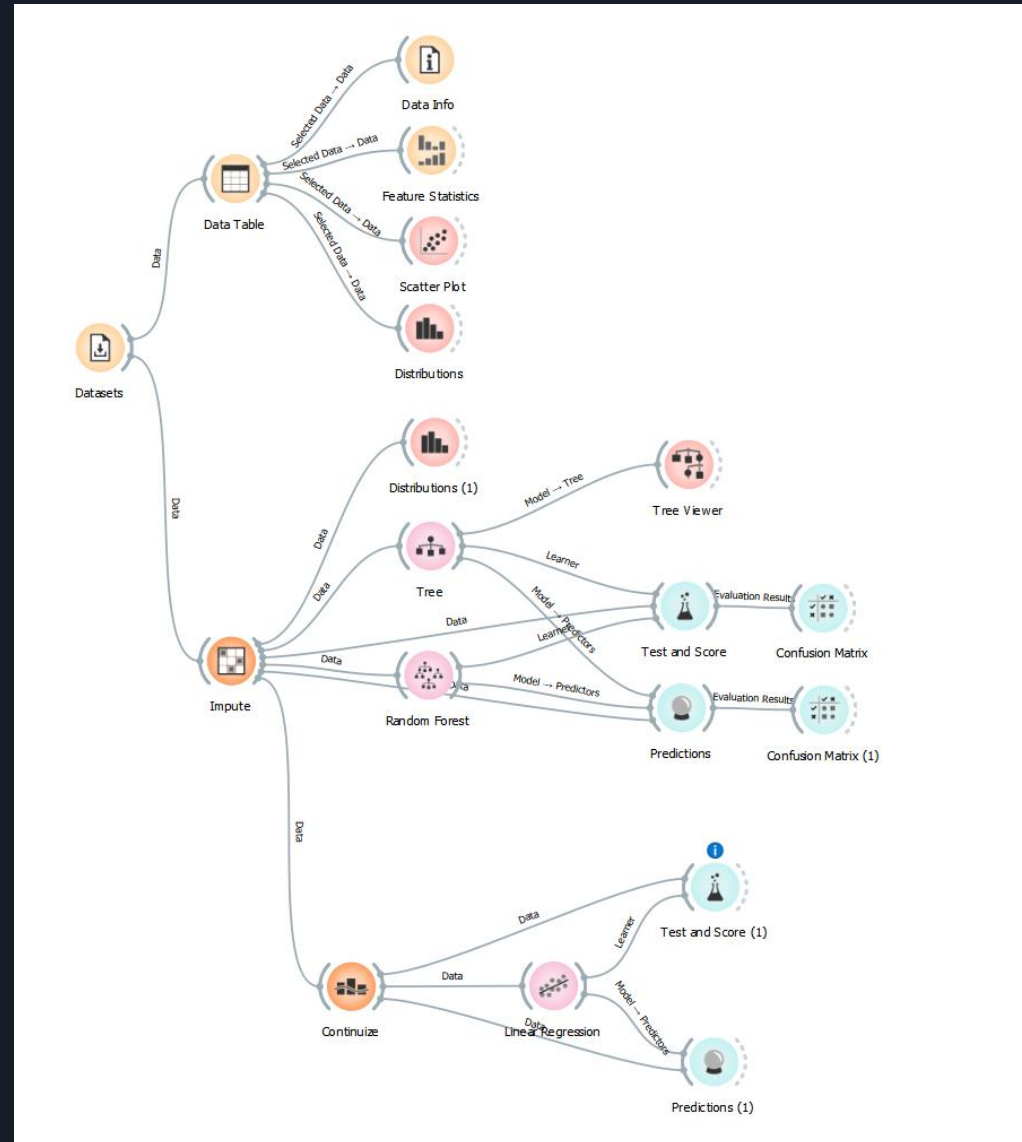
Terlihat dari visualisasi program yang dikerjakan. Dataset diolah menjadi data table yang dianalisis dan dilihat data info, dan statistik dari data yang dipakai, yang Selanjutnya divisualisasikan.





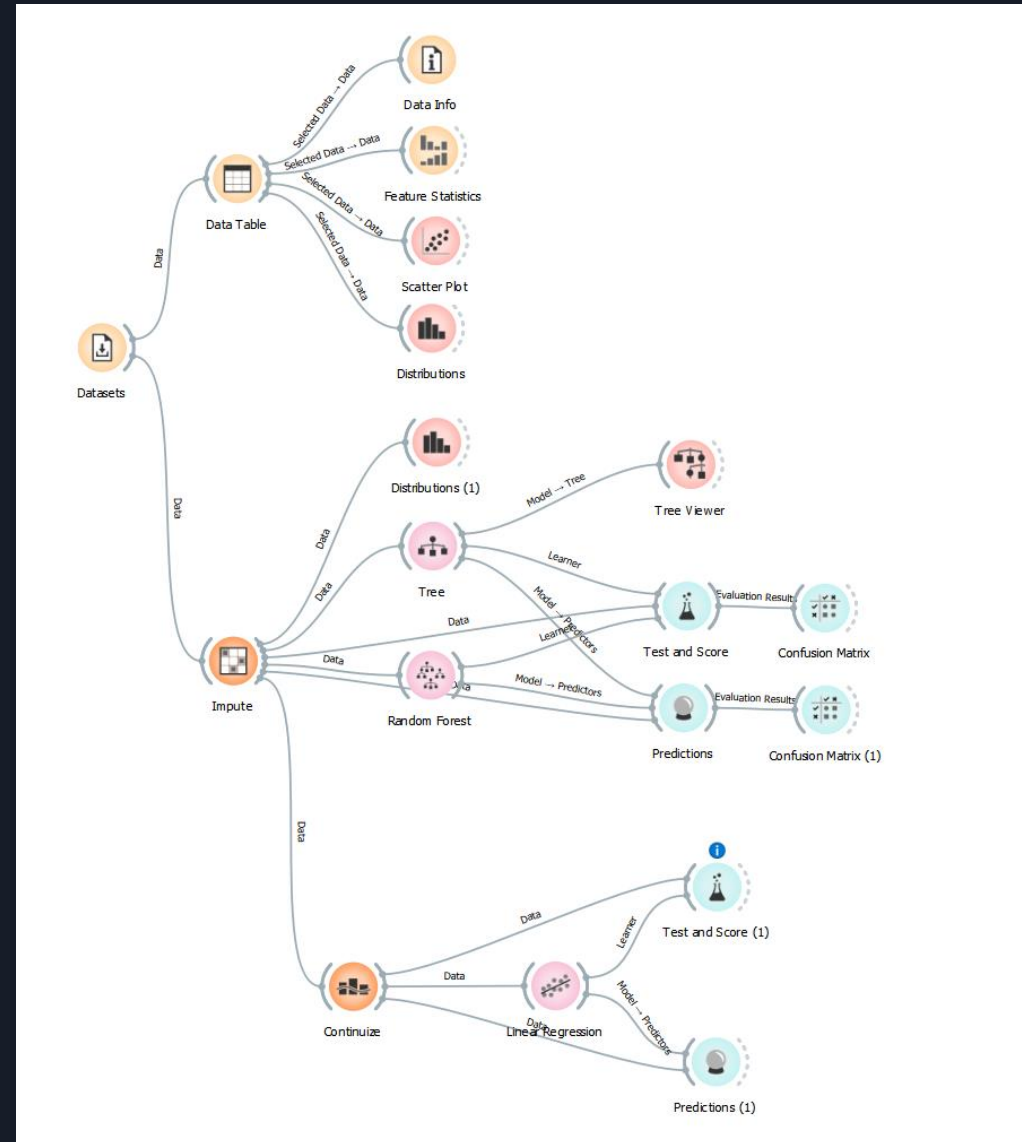
# RESULT

Selanjutnya dalam cabang impute, menghilangkan nilai kosong dalam data supaya model dapat bekerja dengan baik.



# RESULT

pada model linear menggunakan cabang sendiri dikarenakan linear regression hanya dapat mengolah data kolom data interger. Dikarenakan hal itu merubah kolom type menjadi interger.



# RESULT

---

- Terlihat dalam model decision tree dan random forest memiliki nilai F1 tinggi, yang menandakan akurasi model tinggi terhadap data.
- Dalam model linear regression terlihat nilai MSE, RMSE, dan MAE rendah, hal ini menandakan nilai error rendah.

Test and Score - Orange

☒ Cross validation  
Number of folds: 5  
☒ Stratified  
☐ Cross validation by feature

Evaluation results for target (None, show average over classes)

Model	AUC	CA	F1	Precision	Recall
Tree	0.940	0.941	0.941	0.941	0.941
Random Forest	0.987	0.960	0.961	0.961	0.960

Test and Score (1) - Orange

☒ Cross validation  
Number of folds: 5

Model	MSE	RMSE	MAE	R2
Linear Regression	0.039	0.197	0.129	0.830

# RESULT

---

Terlihat contoh chatGPT menjawab pertanyaan yang dijawab dengan cepat.

A

linear regression error Numeric class variable expected.



Linear regression is a type of supervised machine learning algorithm used for predicting numerical (continuous) values based on input features. In most implementations of linear regression, the target variable (also known as the dependent variable or response variable) is expected to be a continuous numerical variable, and the input features (also known as independent variables or predictors) are also expected to be numeric.



The error message "Numeric class variable expected" typically occurs when you are trying to use linear regression with a target variable that is not numeric, but rather categorical or discrete. Linear regression requires a continuous target variable to estimate the relationship between the input features and the target variable using a linear model.

To resolve this error, you can check the data type of your target variable and make sure it is a numeric (continuous) variable. If it is not numeric, you may need to convert it to a numeric format or choose a different algorithm that is suitable for categorical or discrete variables, such as logistic regression for binary classification or decision trees for multi-class classification. Alternatively, if your target variable represents an ordinal variable with a meaningful order, you could consider using ordinal regression techniques instead of linear regression.