MEE 3033: KEPINTARAN BUATAN / *ARTIFICIAL INTELLIGENCES*

SESI PENGAJIAN: SEM 2 (2021/2022)

KUMPULAN KULIAH MEE 3033 (A211): A

(*INDIVIDUAL ASSIGMENT*)

Auto Model – Classification (RapidMiner)

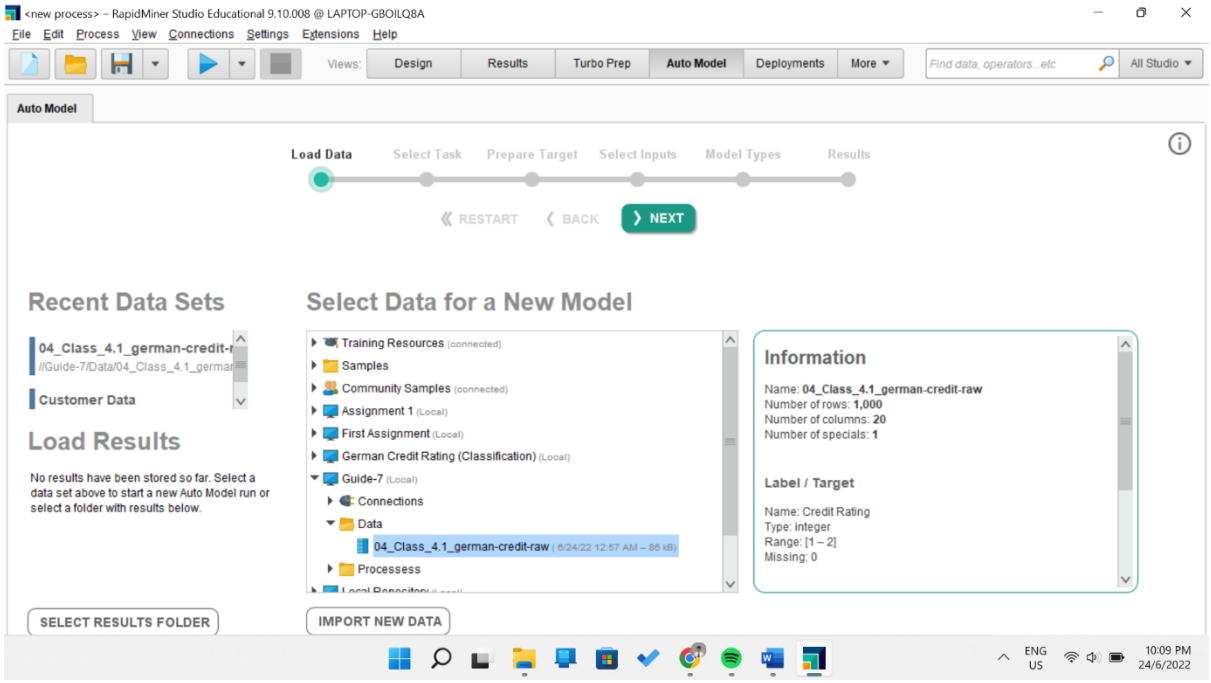
NAMA	NO MATRIC	NO. HP	PROGRAM
MOHD IZZUL IKHWAN BIN MOHD YUSOF	D20201095609	0197818481	AT20 – IJAZAH SARJANA MUDA PENDIDIKAN (TEKNOLOGI MAKLUMAT)

NAMA PENSYARAH: PROFESOR MADYA DR. BAHBIBI BINTI RAHMATULLAH

Auto Model – Classification

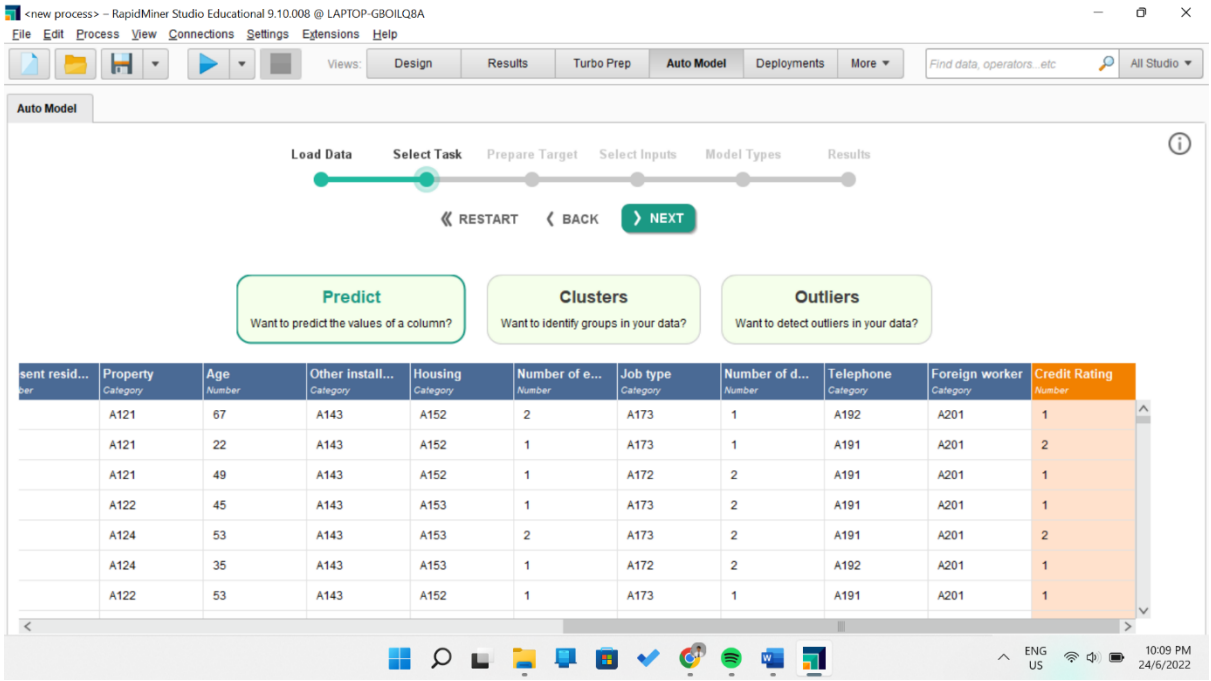
In Auto Model, they have a feature that call auto model features that provides all tools that suitable for beginner or newbies.

01. Select Data



You can choose between importing new data from your own computer or choosing a data from the RapidMiner repository itself.

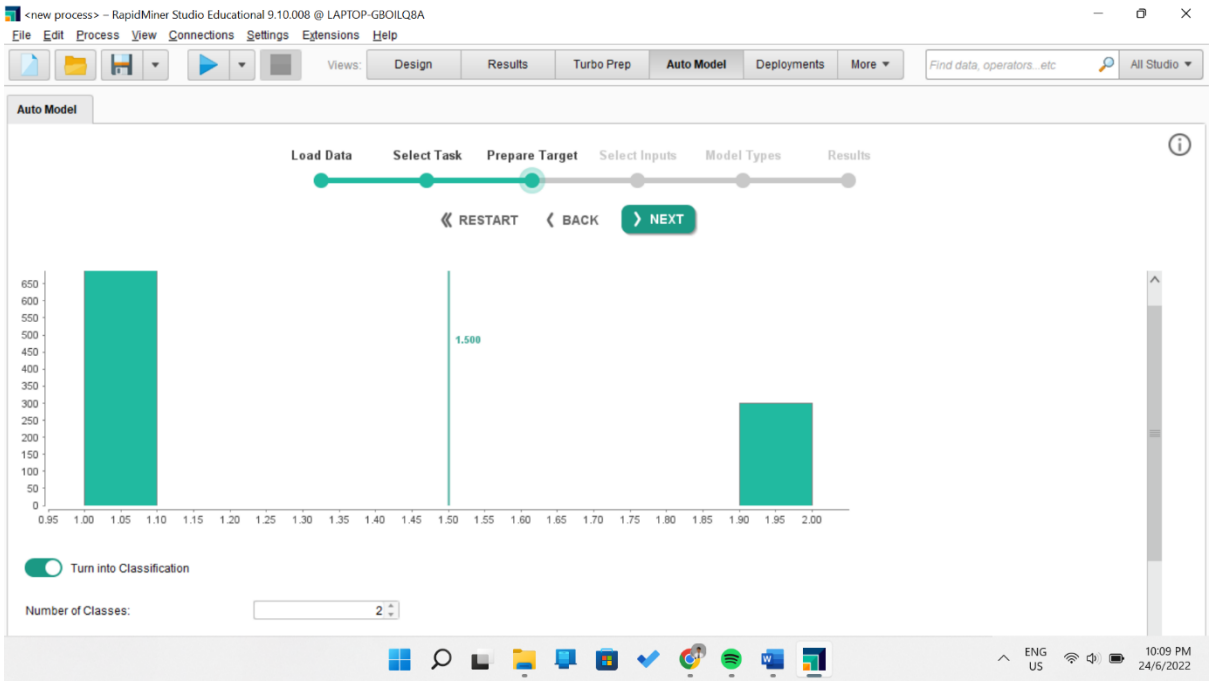
02. Select Task



sent resid...	Property	Age	Other install...	Housing	Number of e...	Job type	Number of d...	Telephone	Foreign worker	Credit Rating
	A121	67	A143	A152	2	A173	1	A192	A201	1
	A121	22	A143	A152	1	A173	1	A191	A201	2
	A121	49	A143	A152	1	A172	2	A191	A201	1
	A122	45	A143	A153	1	A173	2	A191	A201	1
	A124	53	A143	A153	2	A173	2	A191	A201	2
	A124	35	A143	A153	1	A172	2	A192	A201	1
	A122	53	A143	A152	1	A173	1	A191	A201	1

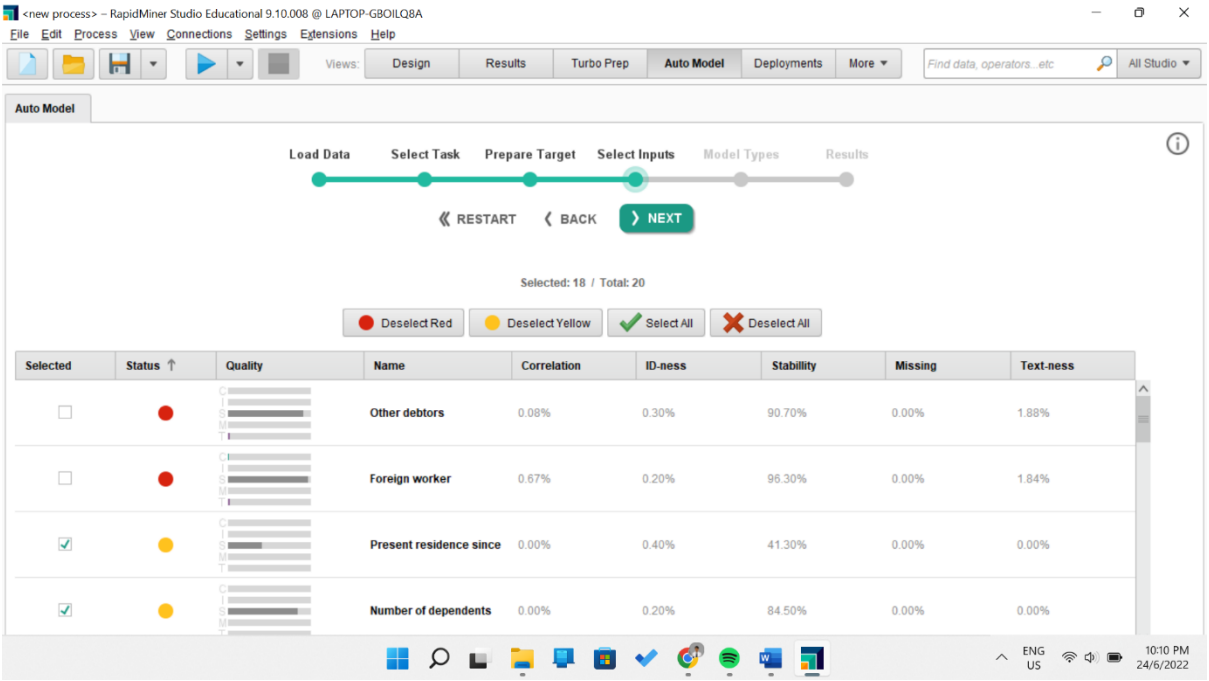
The table displays all the qualities and information. Select which traits or which column you wish to forecast, then select to do so. (Amendment: "label" is given the role attribute.)

03. Prepare Target



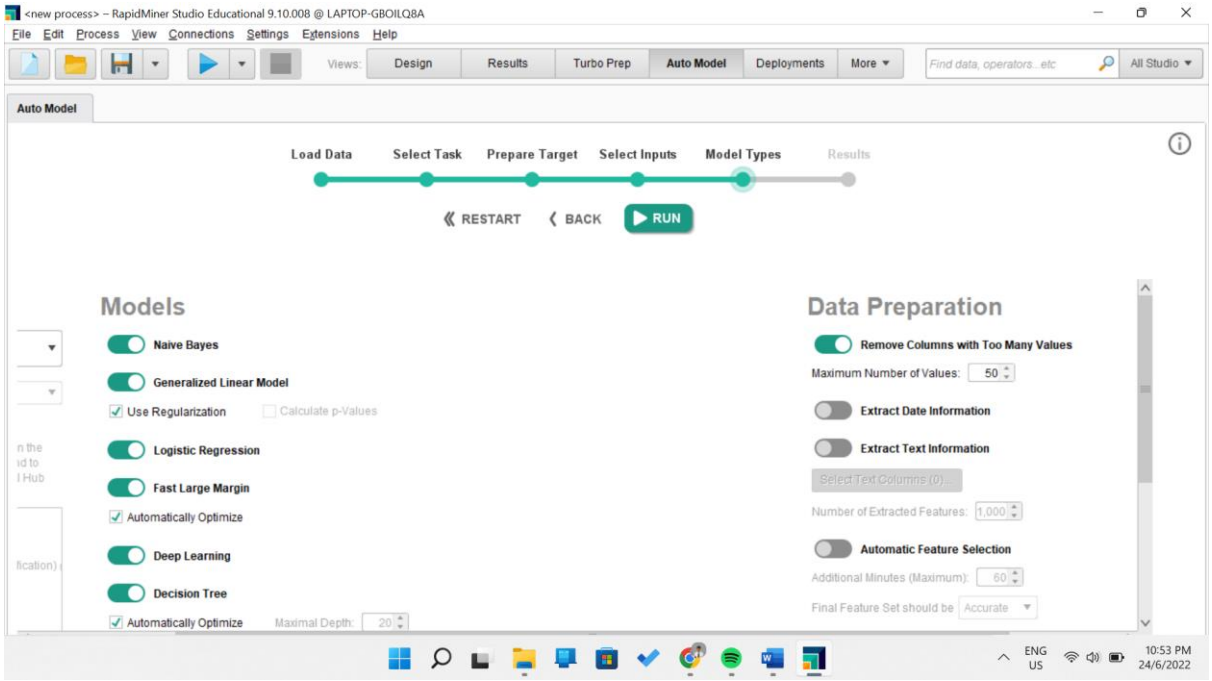
The two separate Credit Rating values are displayed in this green column chart. the classifying process.

04. Select Inputs



A data column's quality is indicated by the coloured status bubble. Red: A red bubble denotes a poor-quality column, which you should usually eliminate from the data collection. A column that acts like an ID but also seems to be text or that has a very low or very high correlation with the target column is indicated by a yellow bubble. Only when the job is marked as "Predicted" may correlation-based yellow bubbles be seen. As you can see, each of those colourful bubbles contains words with meanings beginning with C, I, S, M, and T. Correlation, ID-ness, Stability, Missing, and Text-ness are all capitalised.

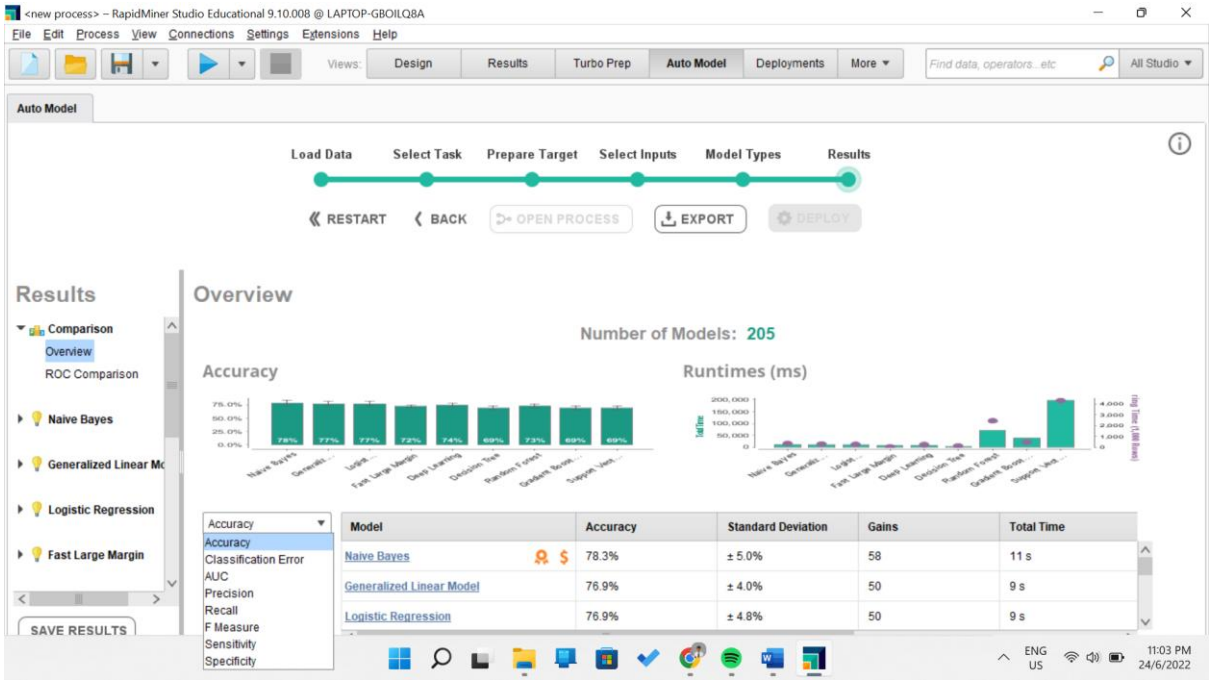
05. Models Type



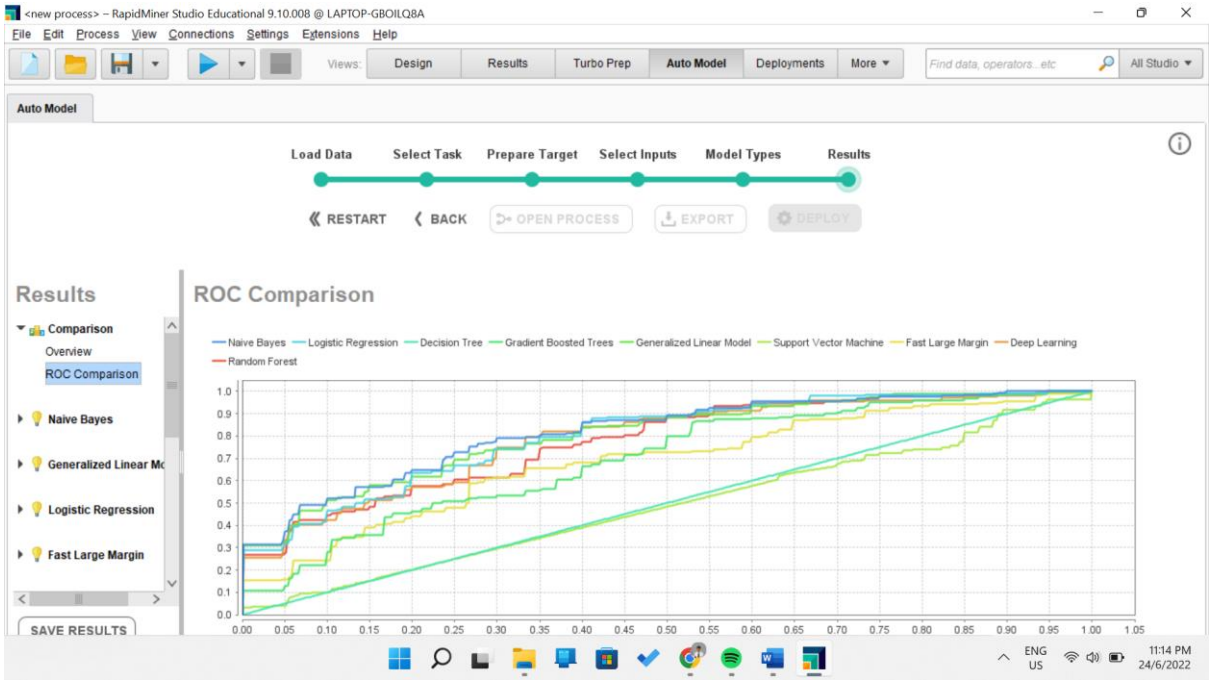
They also offer a tonne of functionality regarding each function in model types. However, I would like to leave it on the default setting.

06. Result

- Comparison



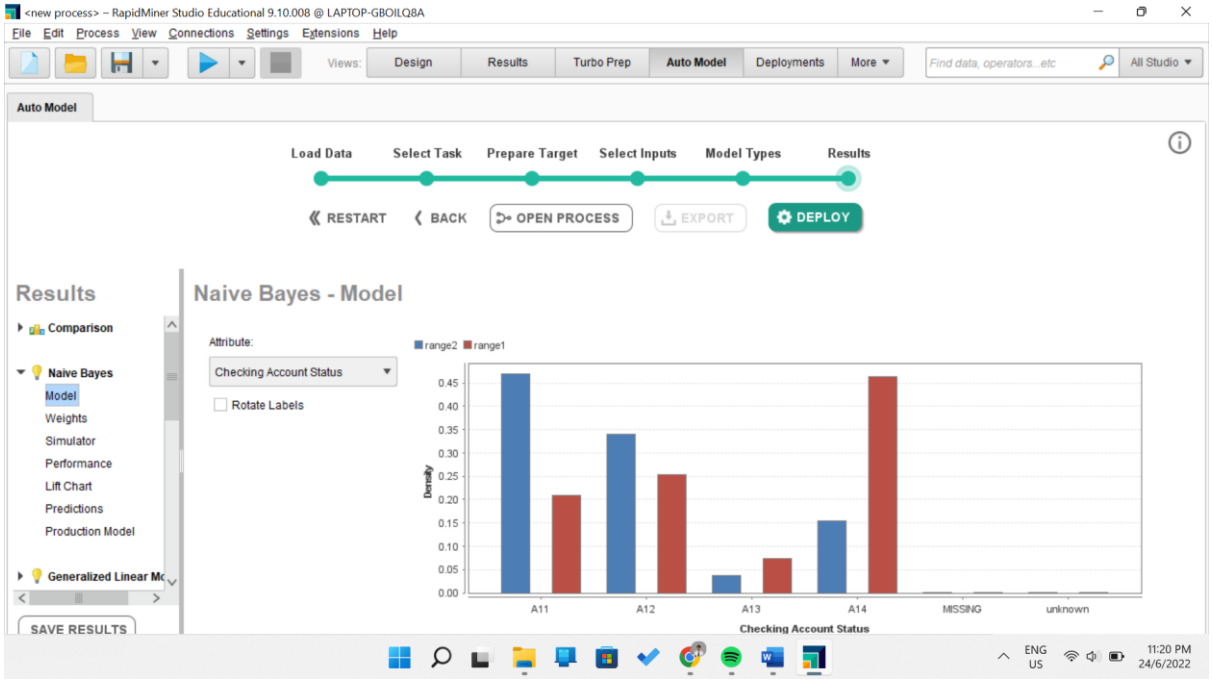
The results were presented in several ways. The first display, also known as the overview, showed us how many models were developed and what percentage applied to all of them.



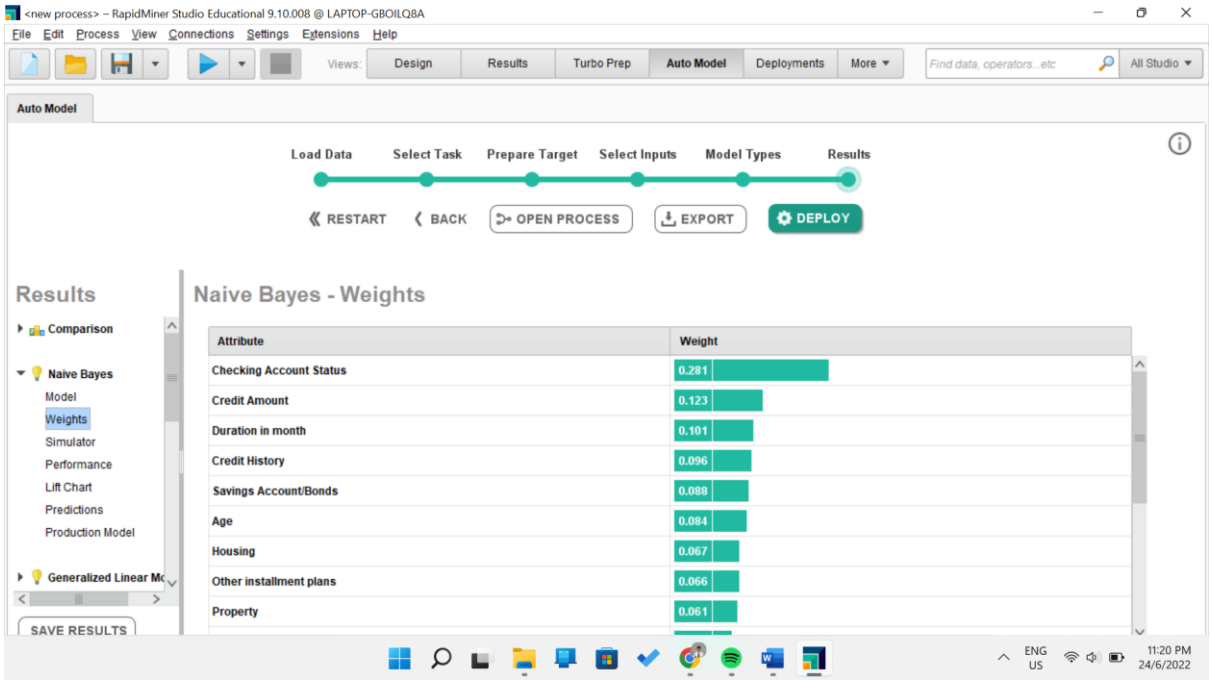
We've also seen ROC comparison in comparison. displays the ROC curves for all models on a single graph. The model is better the closer a curve is near the top left corner. presented only for issues with two classes.

//The findings include a variety of models, such as Naive Bayes, Generalized Linear Model, Logistic Regression, Fast Large Margin, Deep Learning, Decision Tree, and many others. Therefore, I've determined that we should only display 2 or 3 models.//

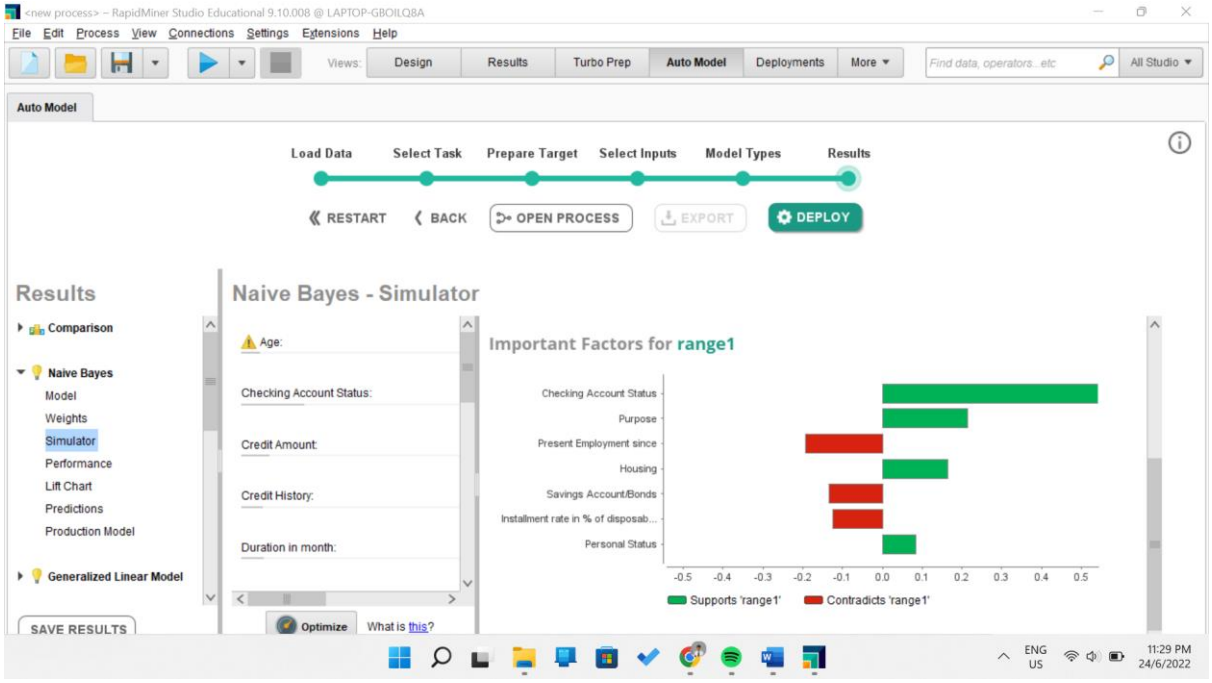
- Naïve Bayes



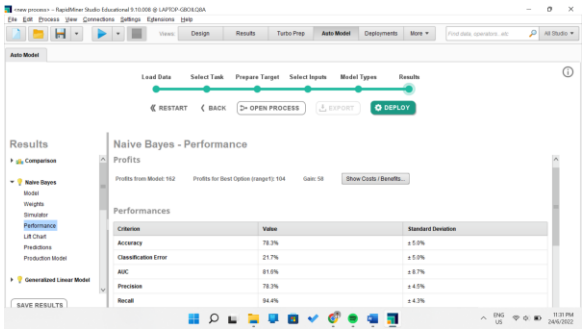
Model: wherever possible, displays a graphical view of the model.

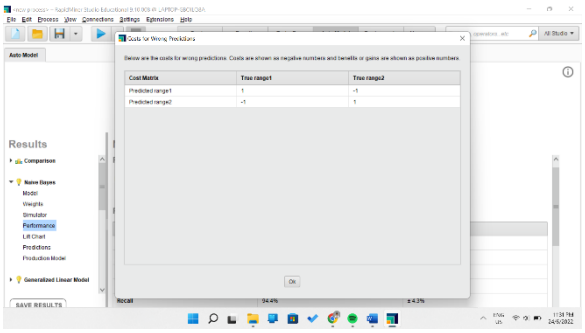


Weights: demonstrates which columns have the greatest overall impact on the model's predictions. The option Explain Predictions must have been chosen on the previous screen for the weights to be displayed.

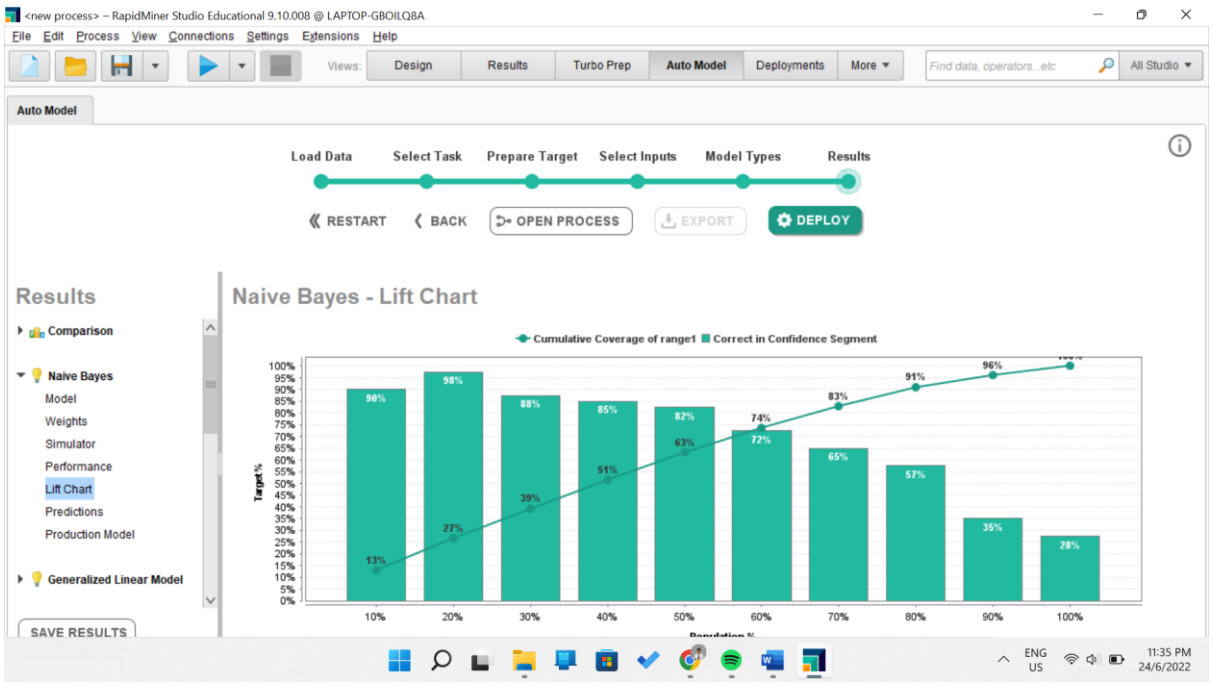


Simulator: enables changing the inputs to a model and viewing the outcome using an intuitive, real-time interface. For those inputs, it displays forecasts, confidences, and justifications.

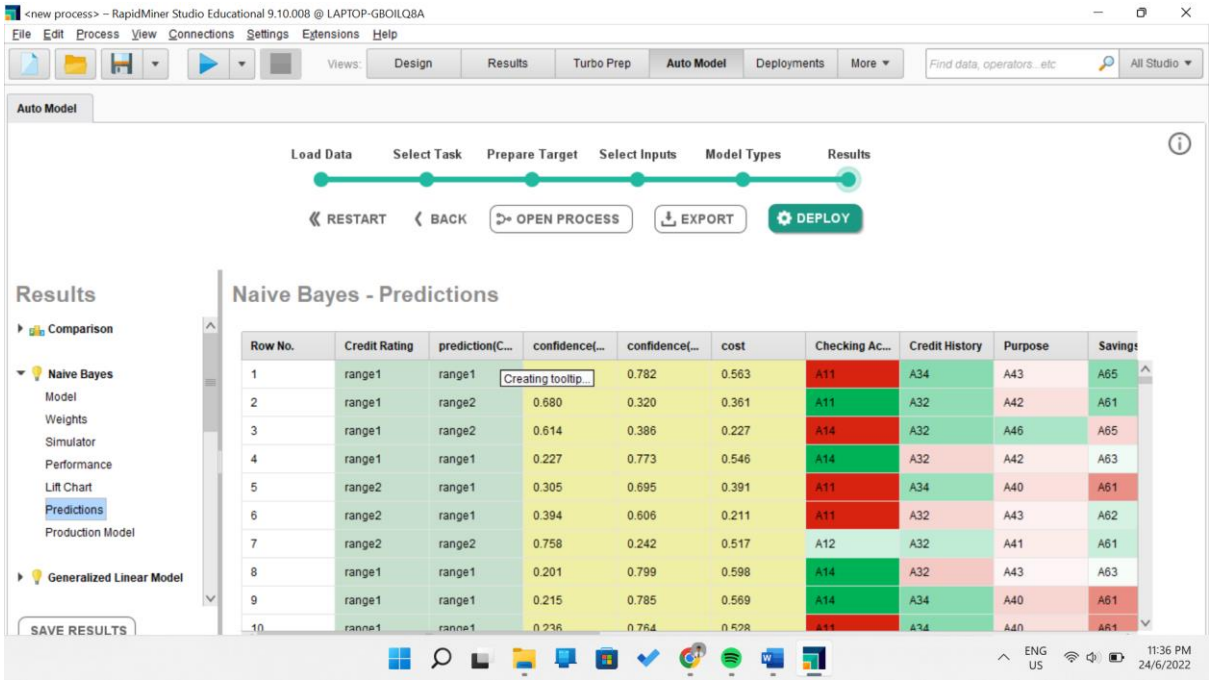




Performance: Depending on the kind of classification problem, provides the model's prediction accuracy and other performance standards. A 40 percent holdout set that hasn't been utilized for any of the model improvements that have been done is used to calculate performance. The performance for seven disjoint subsets is then calculated using this hold-out set as input for a multi-hold-out set validation. The average of the remaining five performances—after the largest and highest performance are eliminated—is presented here. This technique finds a fair compromise between runtime and model validation quality, even if this validation is not as thorough as a full cross-validation.



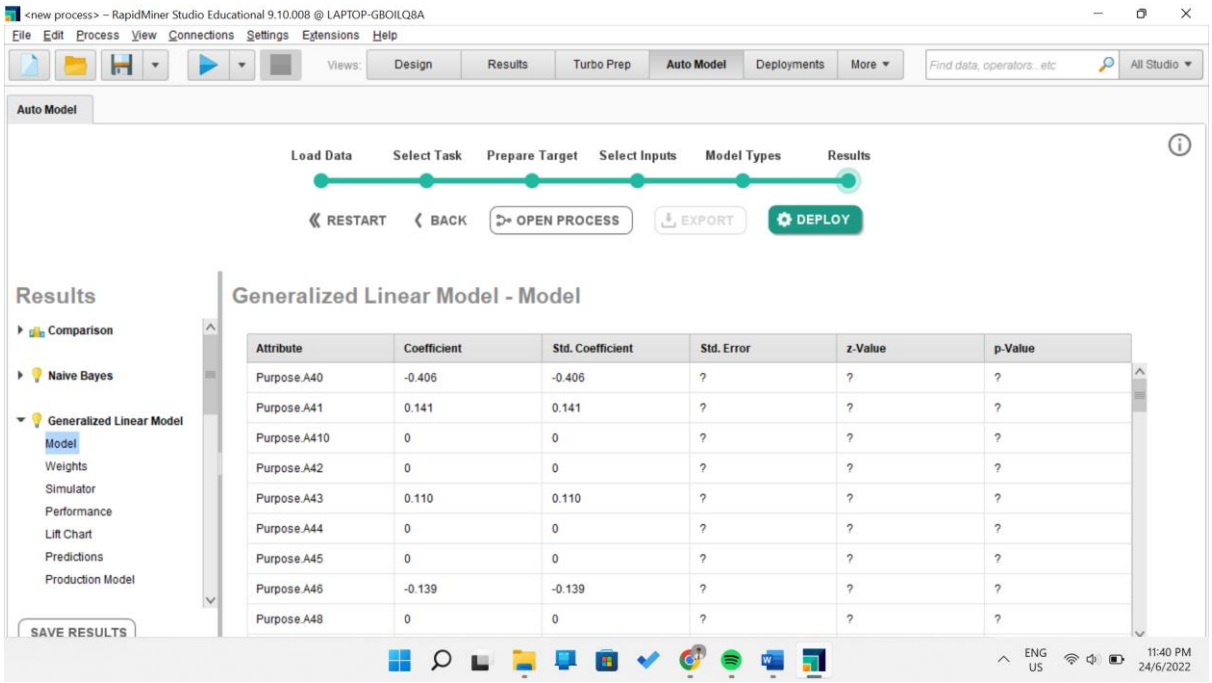
Lift Chart: calculates the ratio between the result achieved with the model and the result obtained without a model to demonstrate the model's efficacy. purely designed for two-class issues.



Row No.	Credit Rating	prediction(C...	confidence(...	confidence(...	cost	Checking Ac...	Credit History	Purpose	Savings
1	range1	range1	Creating tooltip...	0.782	0.563	A11	A34	A43	A65
2	range1	range2	0.680	0.320	0.361	A11	A32	A42	A61
3	range1	range2	0.614	0.386	0.227	A14	A32	A46	A65
4	range1	range1	0.227	0.773	0.546	A14	A32	A42	A63
5	range2	range1	0.305	0.695	0.391	A11	A34	A40	A61
6	range2	range1	0.394	0.606	0.211	A11	A32	A43	A62
7	range2	range2	0.758	0.242	0.517	A12	A32	A41	A61
8	range1	range1	0.201	0.799	0.598	A14	A32	A43	A63
9	range1	range1	0.215	0.785	0.569	A14	A34	A40	A61
10	range1	range1	0.236	0.764	0.528	A11	A34	A40	A61

Predictions: offers a forecast and an explanation for that prediction for the 40 percent hold-out group and each row with a missing value in the target column. Only when the preceding screen's option Explain Predictions was chosen are the explanations displayed.

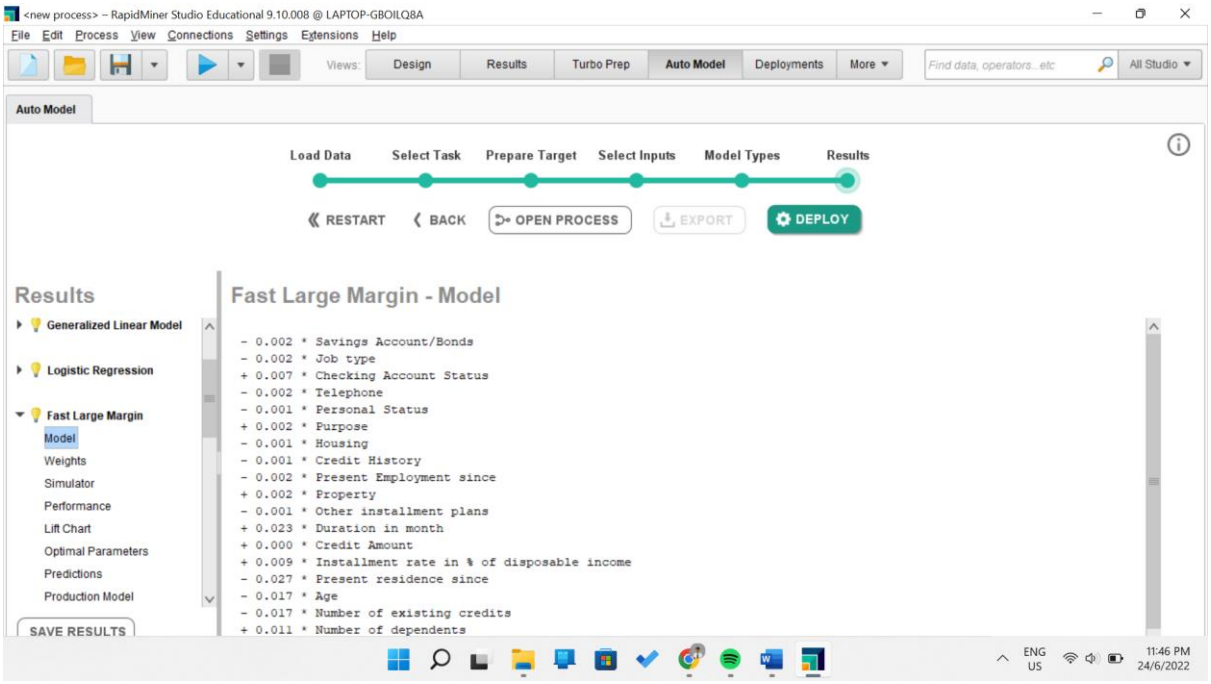
- Generalized Linear Model [model]



Attribute	Coefficient	Std. Coefficient	Std. Error	z-Value	p-Value
Purpose.A40	-0.406	-0.406	?	?	?
Purpose.A41	0.141	0.141	?	?	?
Purpose.A410	0	0	?	?	?
Purpose.A42	0	0	?	?	?
Purpose.A43	0.110	0.110	?	?	?
Purpose.A44	0	0	?	?	?
Purpose.A45	0	0	?	?	?
Purpose.A46	-0.139	-0.139	?	?	?
Purpose.A48	0	0	?	?	?

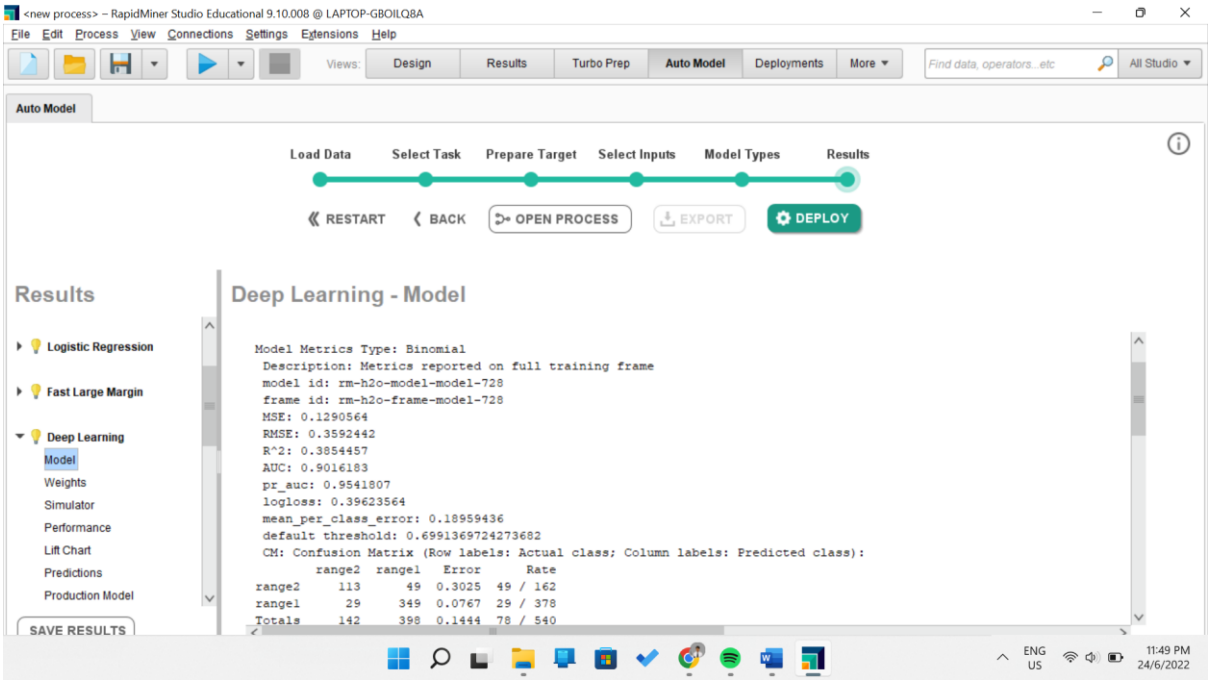
Model: shows a generalized linear model that provide us in table content. The table shows an attribute, coefficient, std.coefficient, std.error, z-value and p-value.

- Fast large Margin [model]



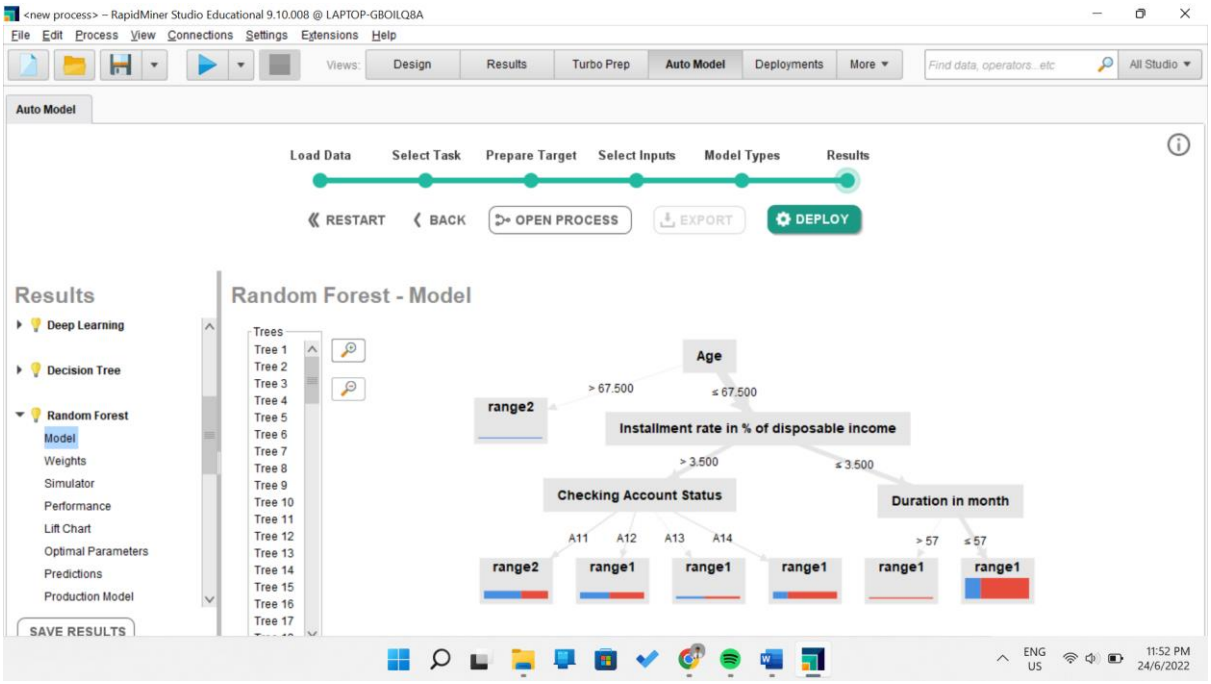
Model: shows a graphical representation of the model, where possible.

- Deep Learning [model]



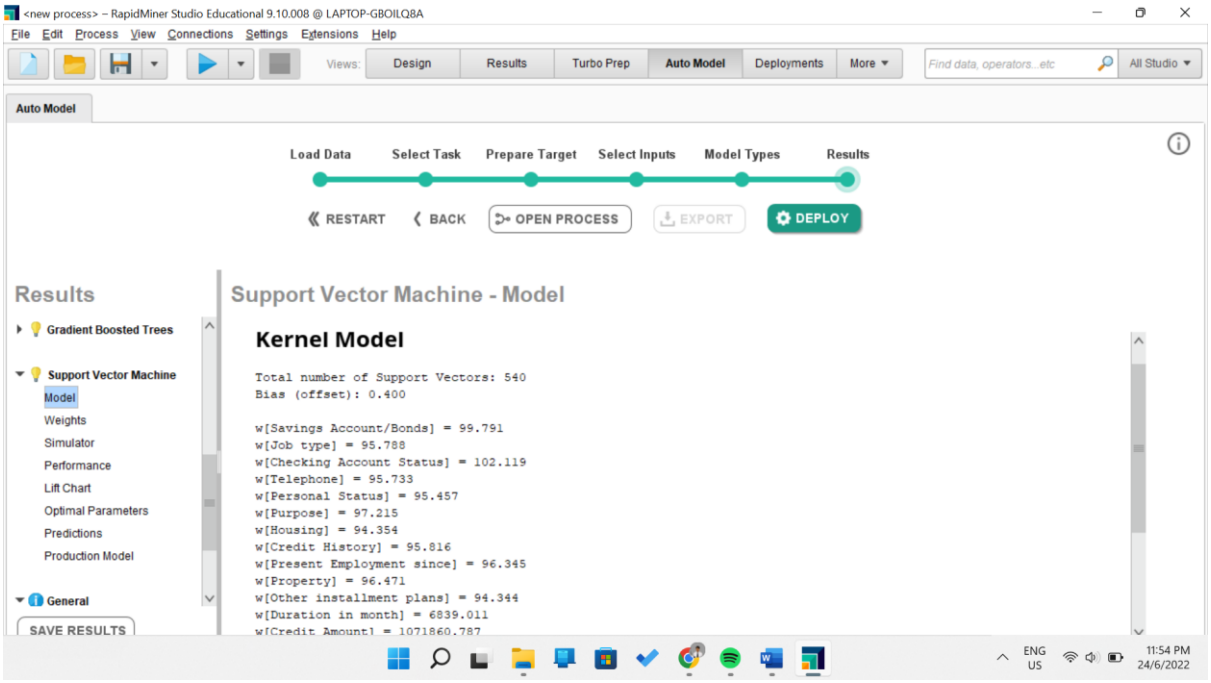
Model: shows a graphical representation of the model. If you can see, our types of model metrics are Binomial.

- Random Forest [model]



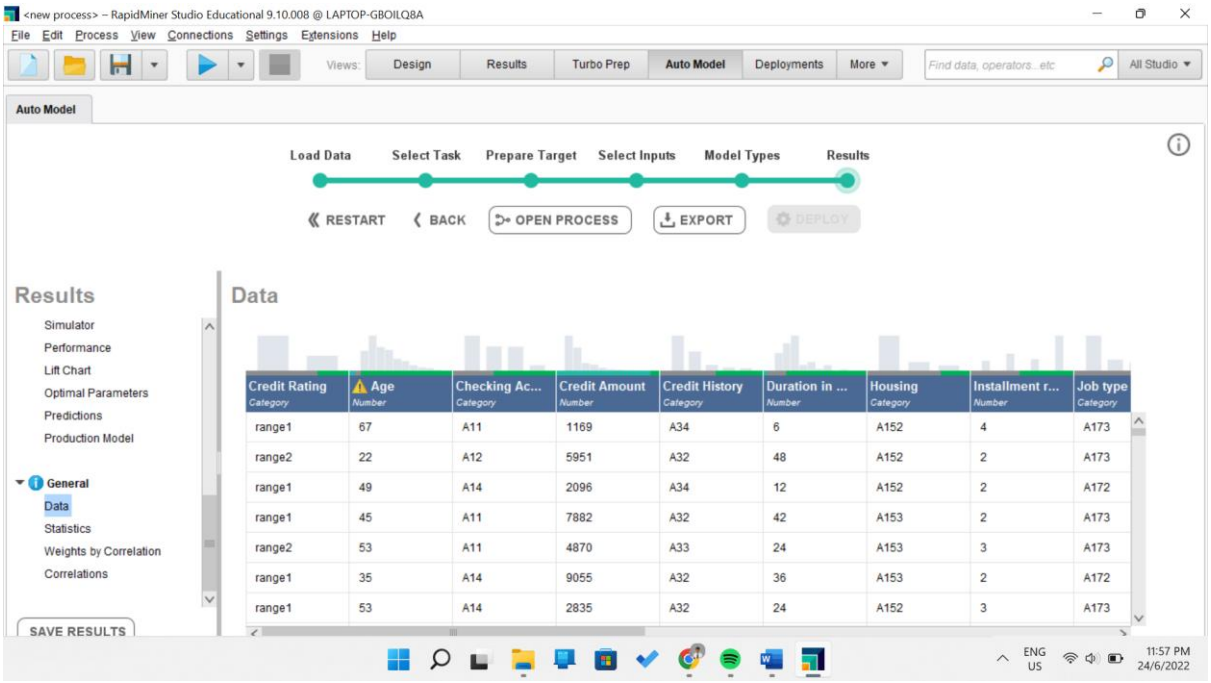
Model: shows a model type decision tree.

- Support Vector Machine [model]

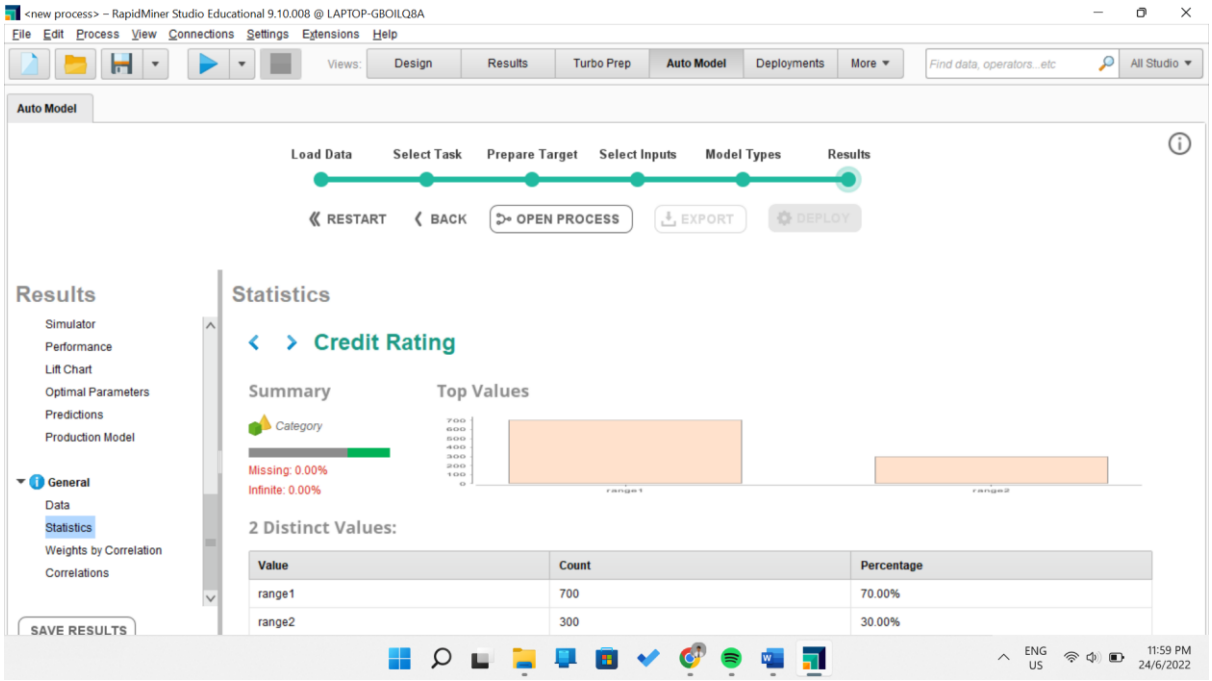


Model: shows a kernel model, that shows total number of support vectors and bias (offset).

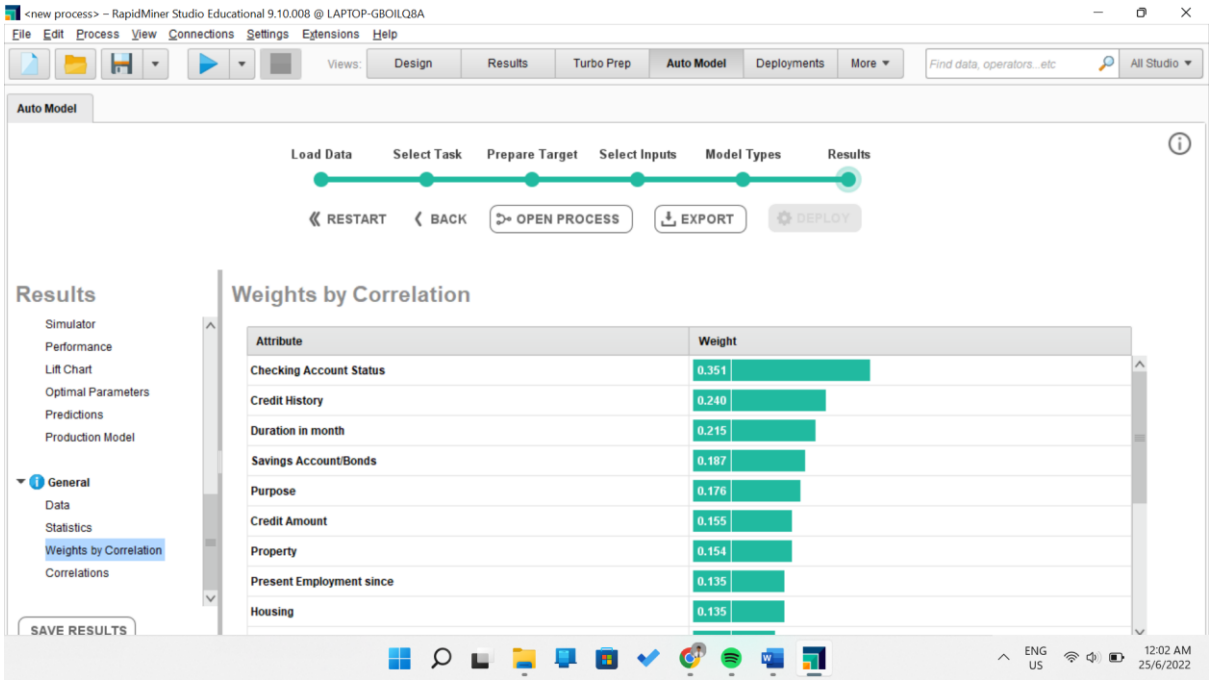
- General



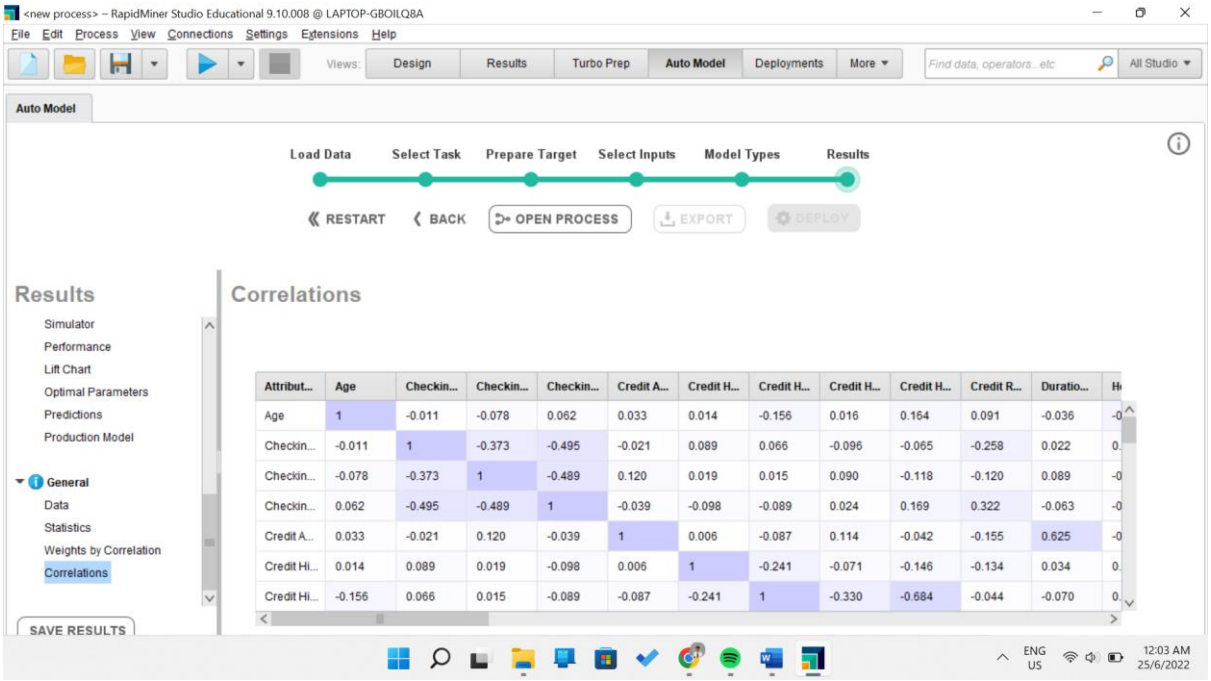
Data: the data collection following its transformation for modelling. This serves as the input for autonomous feature engineering as well as all modelling techniques. Only a portion of this data may be used by the model, and it may also produce new columns.



Statistic: The statistic can show us the graphical data about all the attributes.



Weights by Correlation: regardless of the modelling technique or prospective feature engineering, the overall significance of each of the original input columns for the value of the target column. The correlation between the columns and the predicted column is the basis for the weights. In contrast, the model-specific weights display which columns, on average, have the most impact on each model's predictions.



Correlations: A matrix showing the correlations between columns.