Experiment 03 - Classification(Decision tree and Naïve Bayes' Classification algorithms) using Rapid Miner tool

| Roll No. | 65 |
| --- | --- |
| Name | Shreyansh Singh |
| Class | D15 |
| Subject | Business Intelligence Lab |
| LO Mapped | LO4: Classification using Rapid Miner tool |
|  |  |

**Introduction** :

RapidMiner is a powerful and intuitive data science platform that enables organizations to easily and efficiently extract insights from large volumes of data. It is a popular open-source tool for data mining, machine learning, and predictive analytics.

RapidMiner allows users to easily connect to and integrate data from various sources, such as databases, spreadsheets, and APIs. It also includes a wide range of built-in data transformation and preprocessing tools, which can be used to clean, reshape, and transform data prior to analysis.

One of the key features of RapidMiner is its visual workflow interface, which enables users to build and modify complex data analysis workflows without needing to write code. Users can drag and drop operators and connect them together to create custom data processing and analysis pipelines.

RapidMiner supports a wide range of machine learning algorithms, including supervised and unsupervised learning, as well as deep learning. It also includes tools for model evaluation and optimization, enabling users to fine-tune their models for maximum accuracy.

Overall, RapidMiner is a comprehensive and flexible data science platform that can be used for a wide range of applications, including predictive analytics, customer segmentation, fraud detection, and more.

**Classification**:

Classification algorithms are a type of machine learning algorithm used to predict the class or category of a given data point based on its features or attributes. The goal of classification is to learn a function that maps input features to output classes, so that the function can be used to make predictions on new, unseen data.

There are several types of classification algorithms, including:

Logistic Regression: A statistical algorithm that models the probability of a binary outcome based on one or more predictor variables.

Decision Trees: A tree-like model where each internal node represents a test on an attribute, each branch represents the outcome of the test, and each leaf node represents a class label.

Random Forest: An ensemble learning algorithm that constructs a multitude of decision trees and combines their outputs to make a final prediction.

Support Vector Machines (SVM): A discriminative classifier that separates data into classes using a hyperplane that maximizes the margin between the classes.

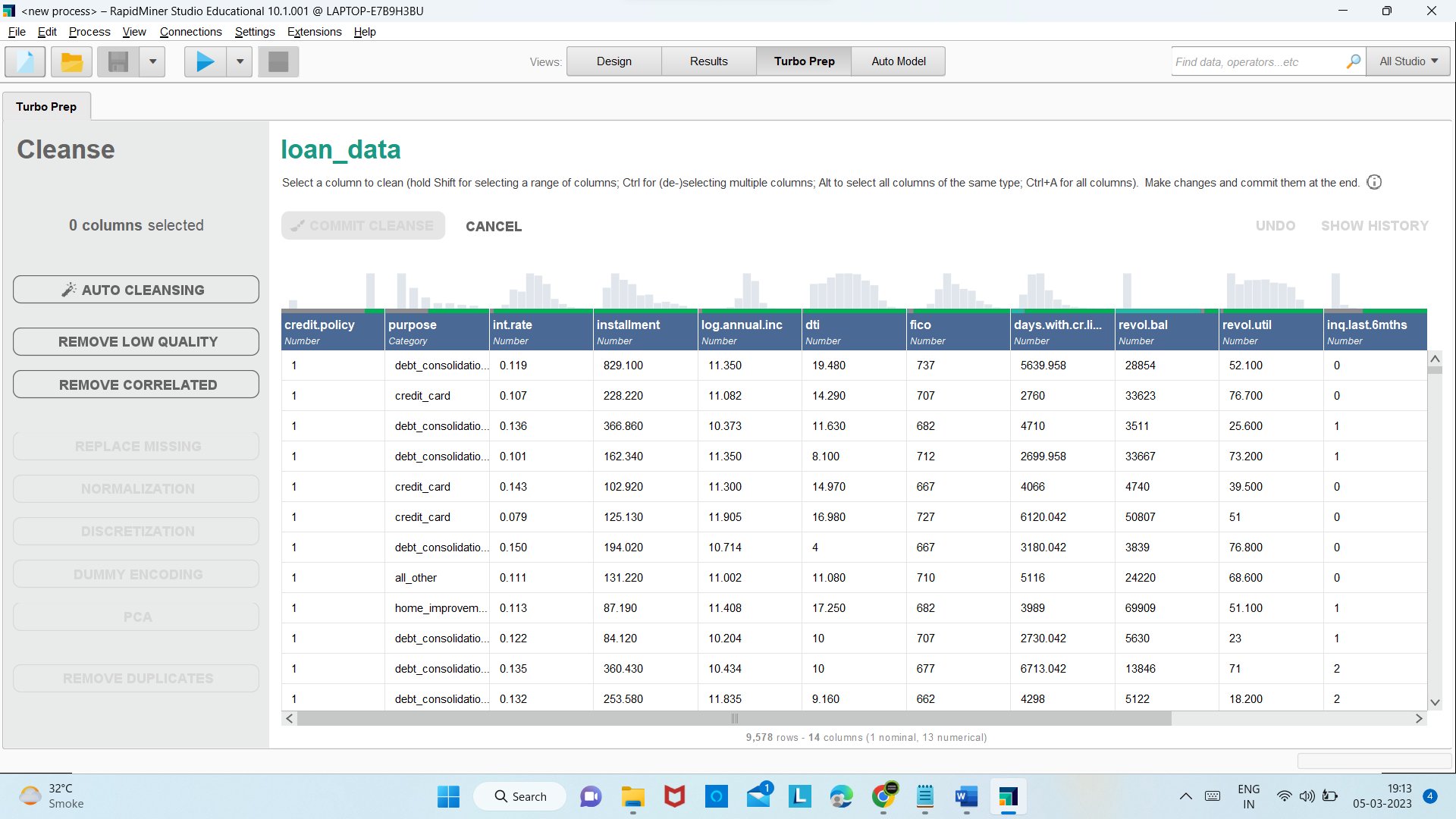
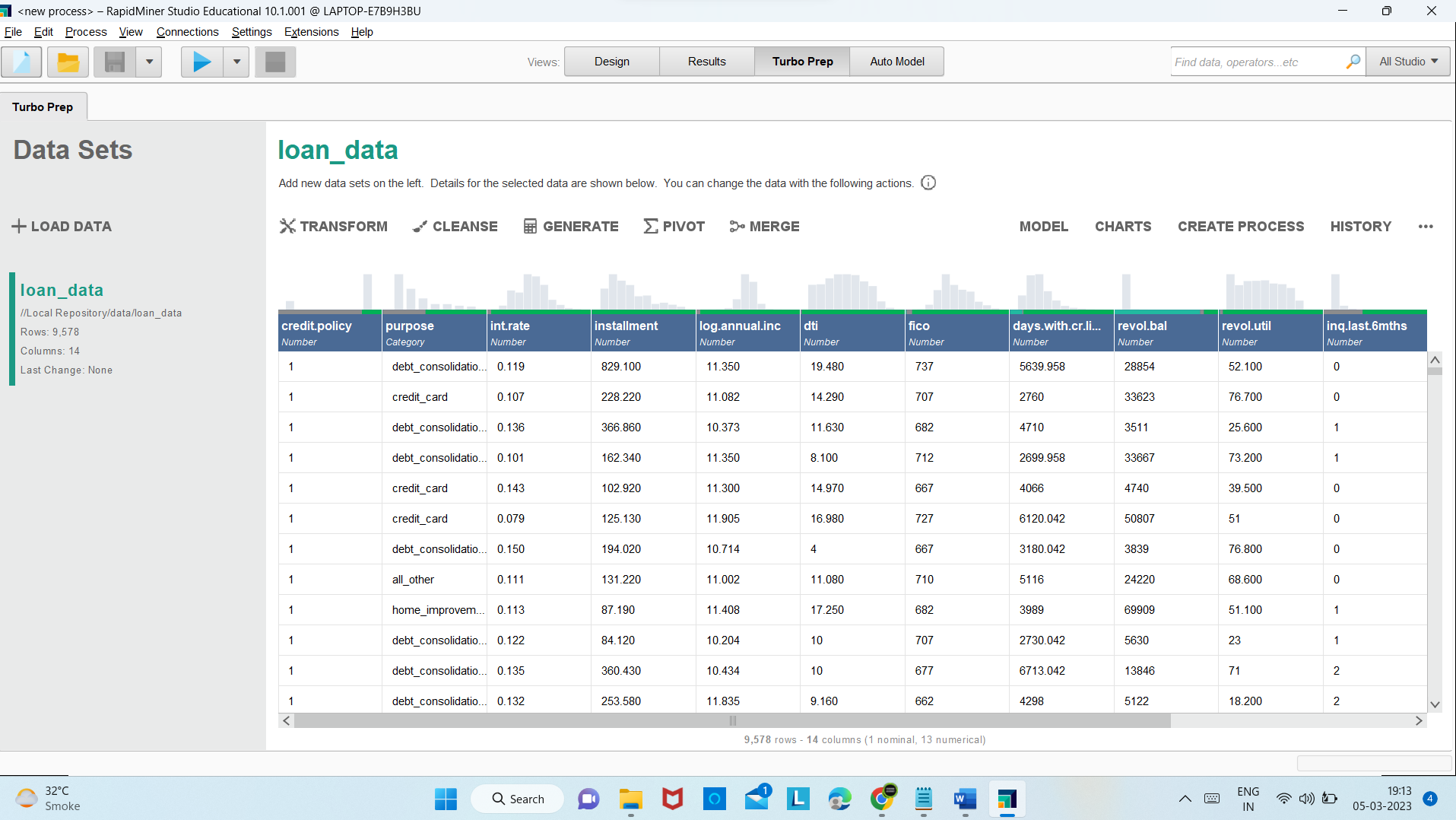
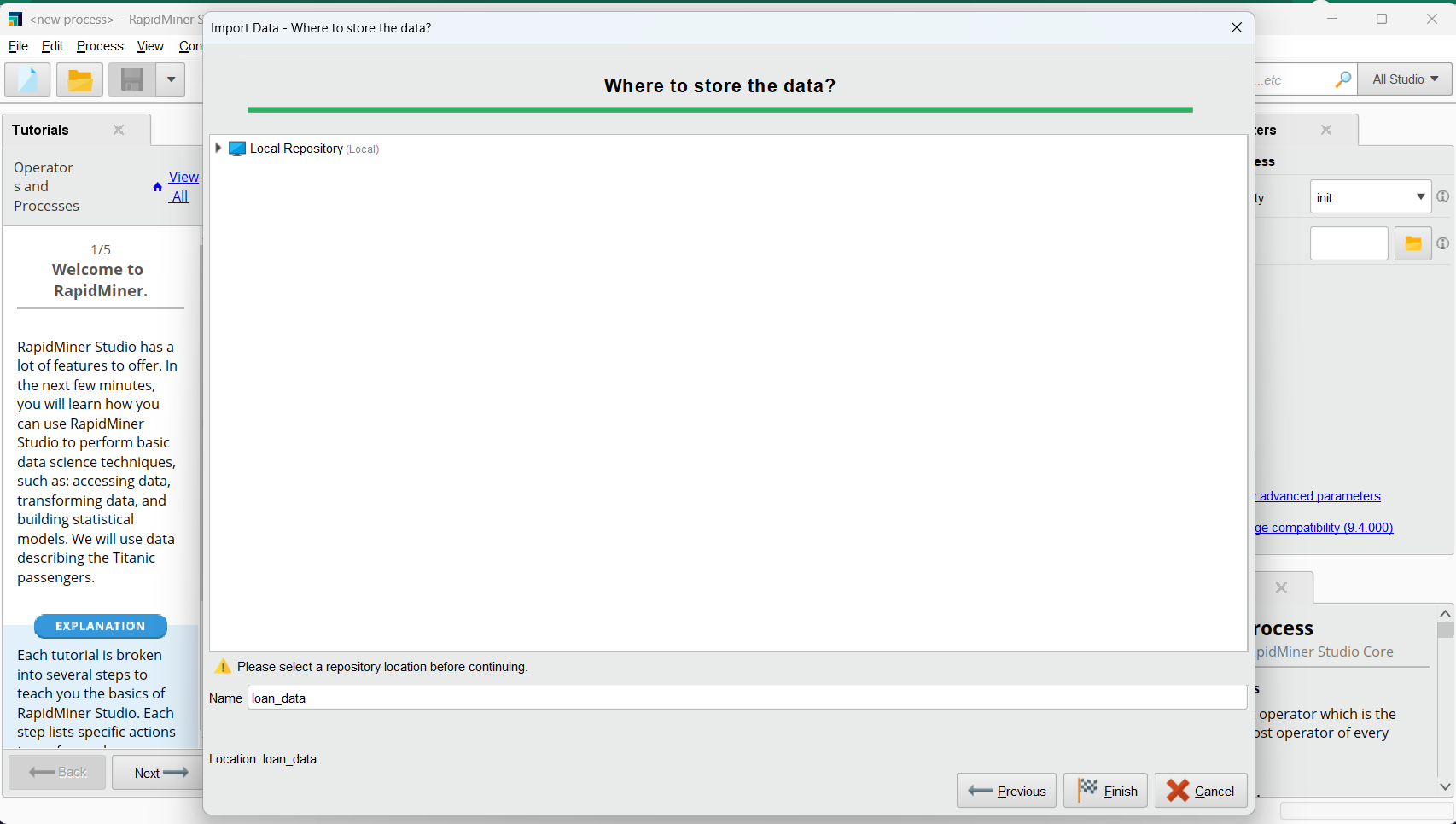
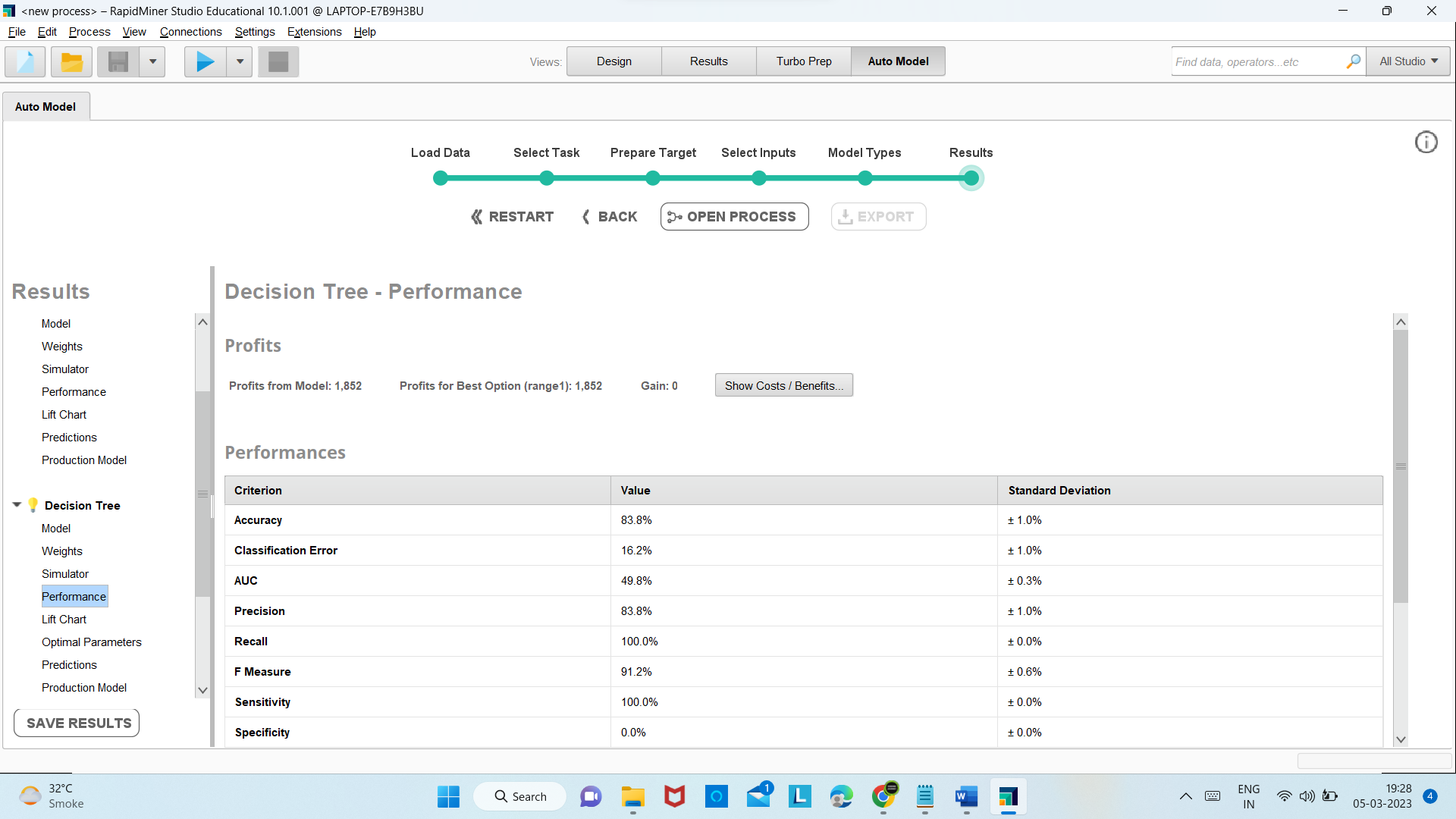
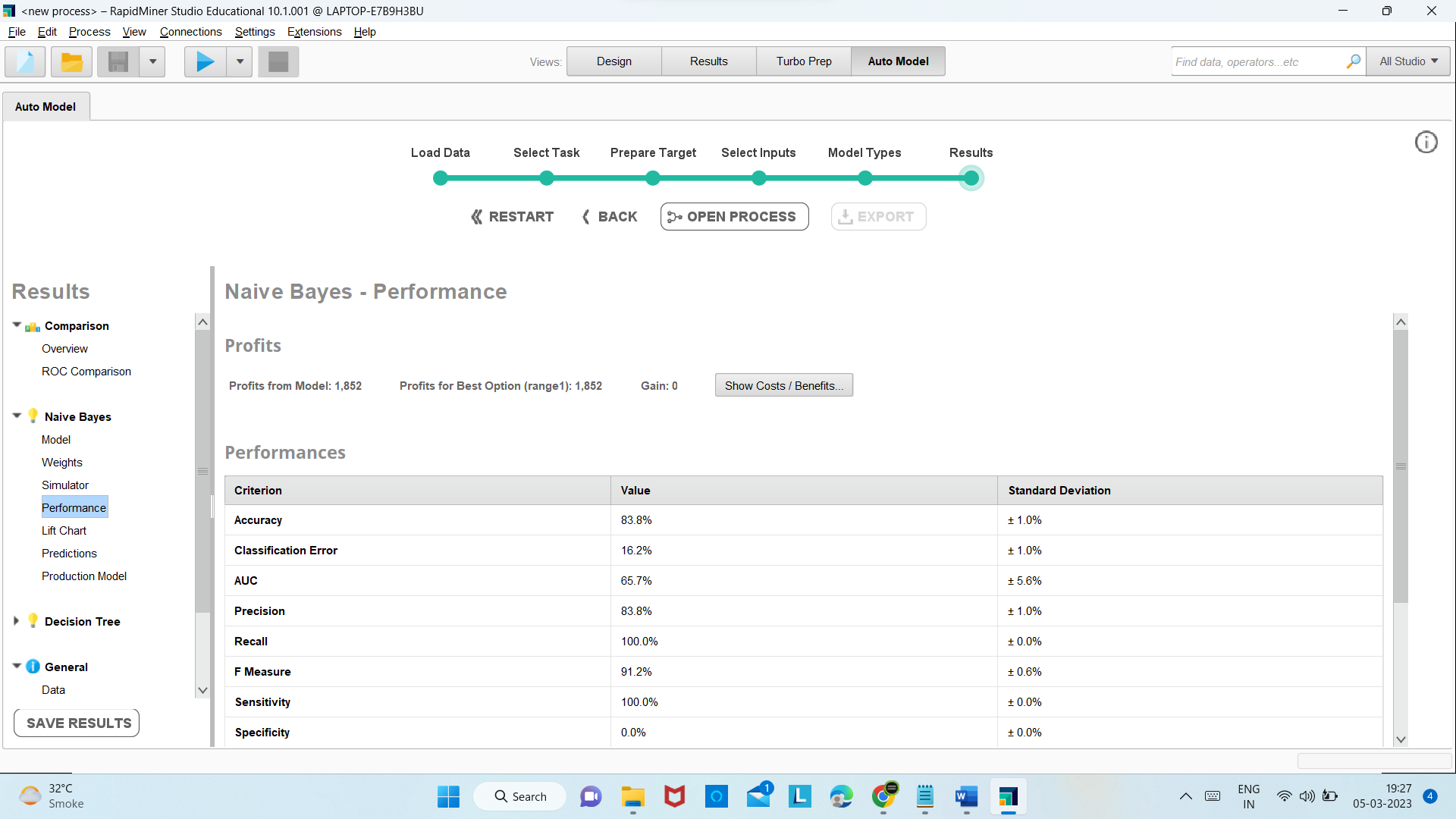
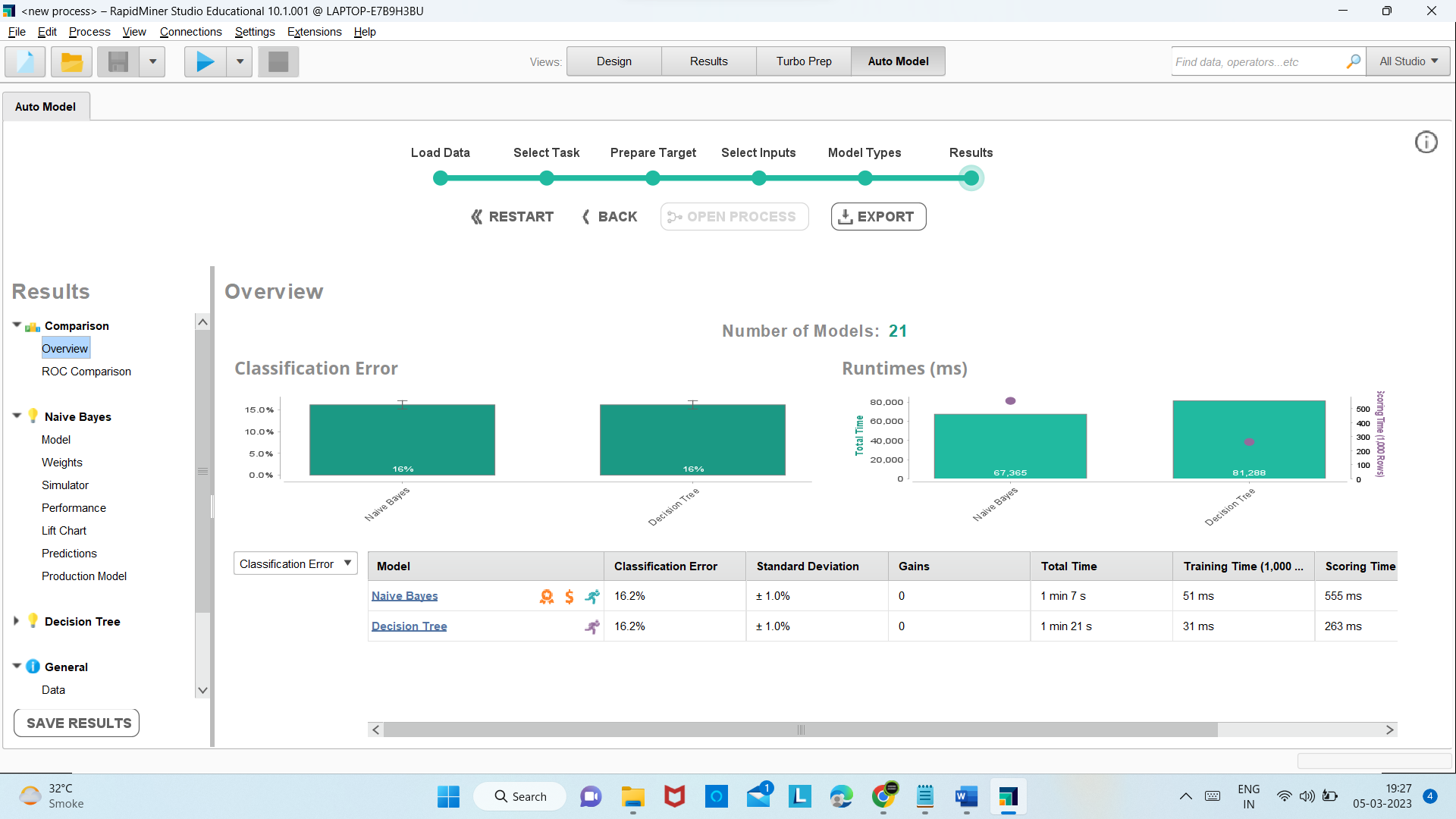
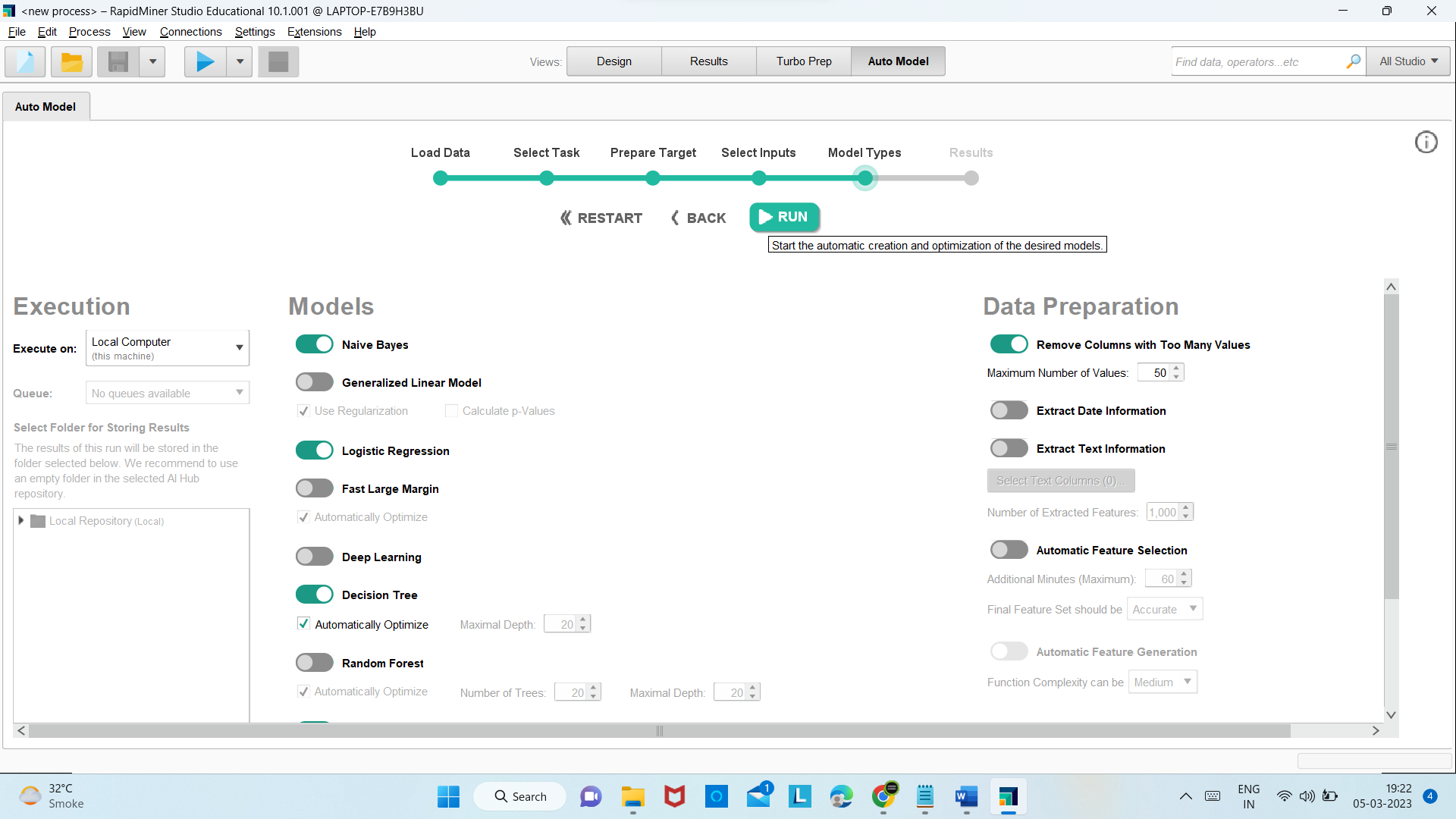
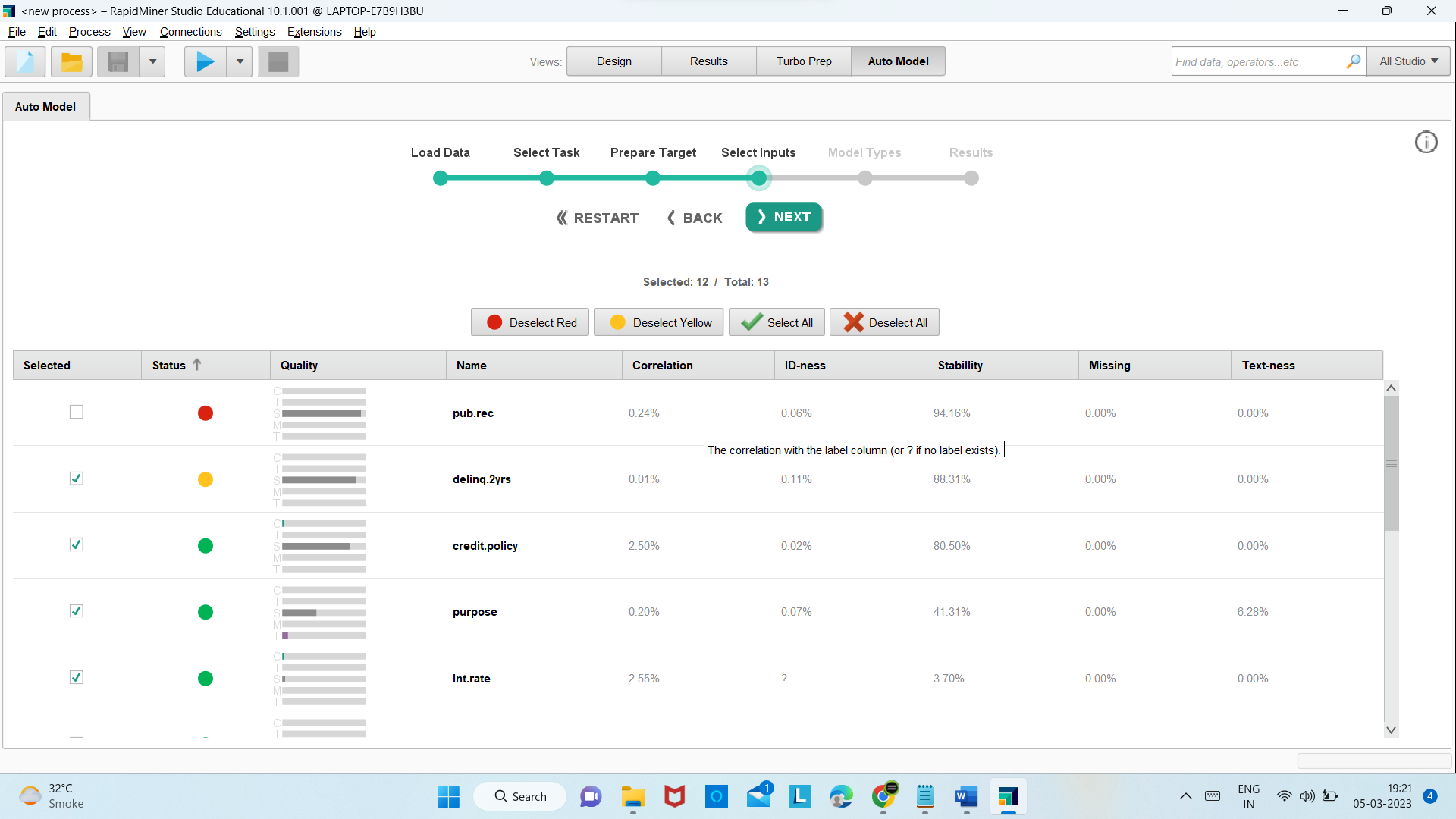
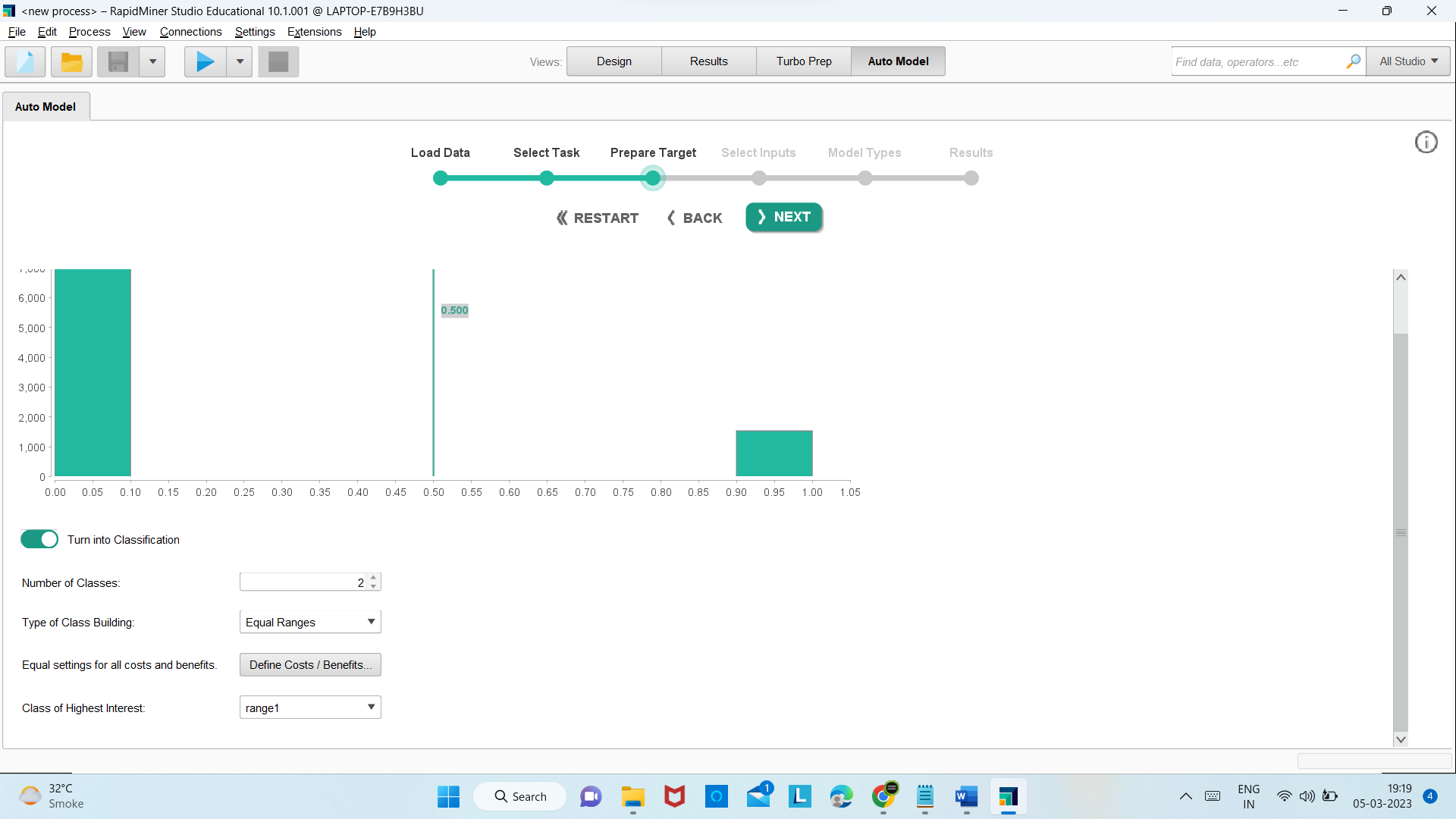
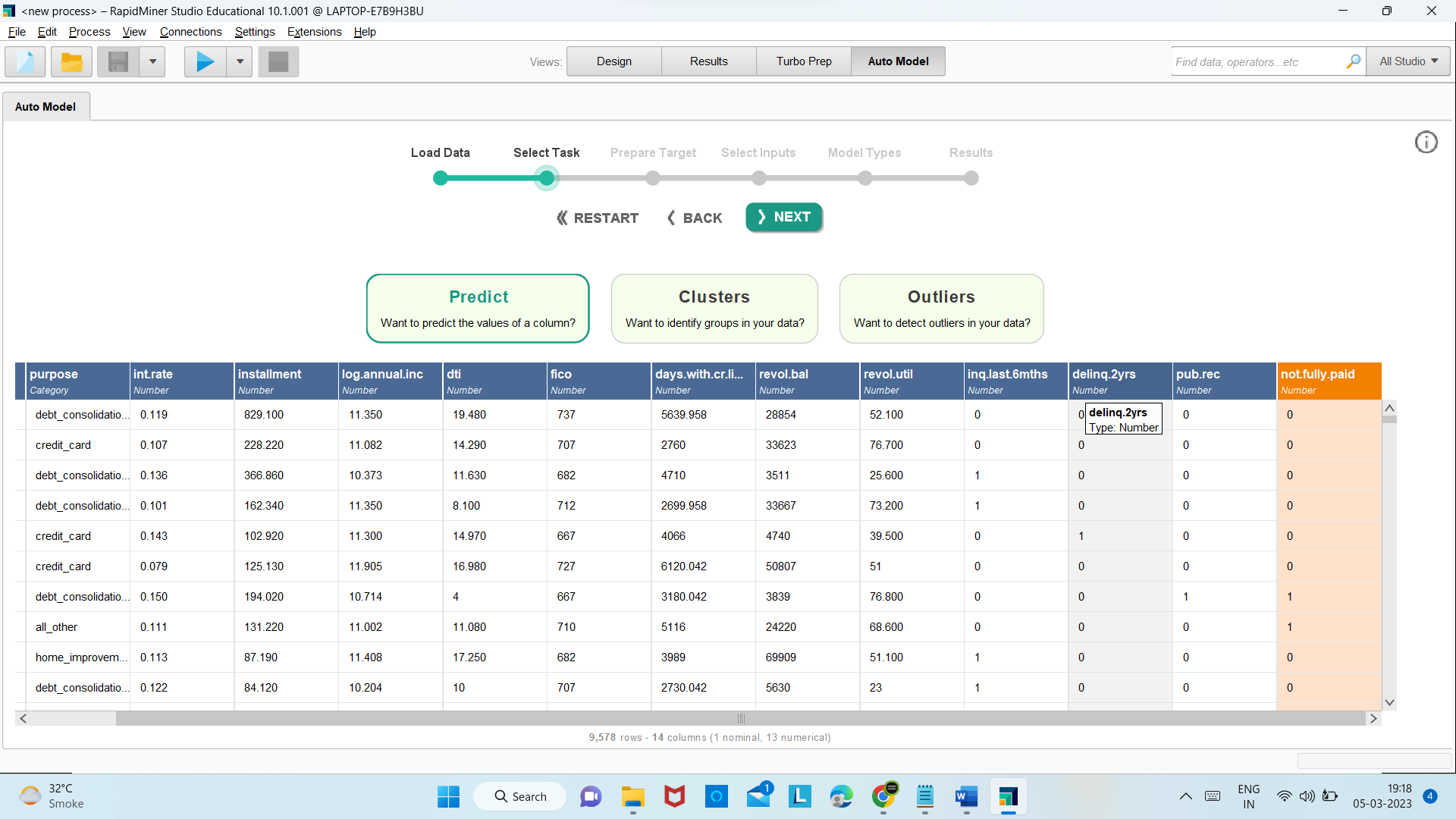
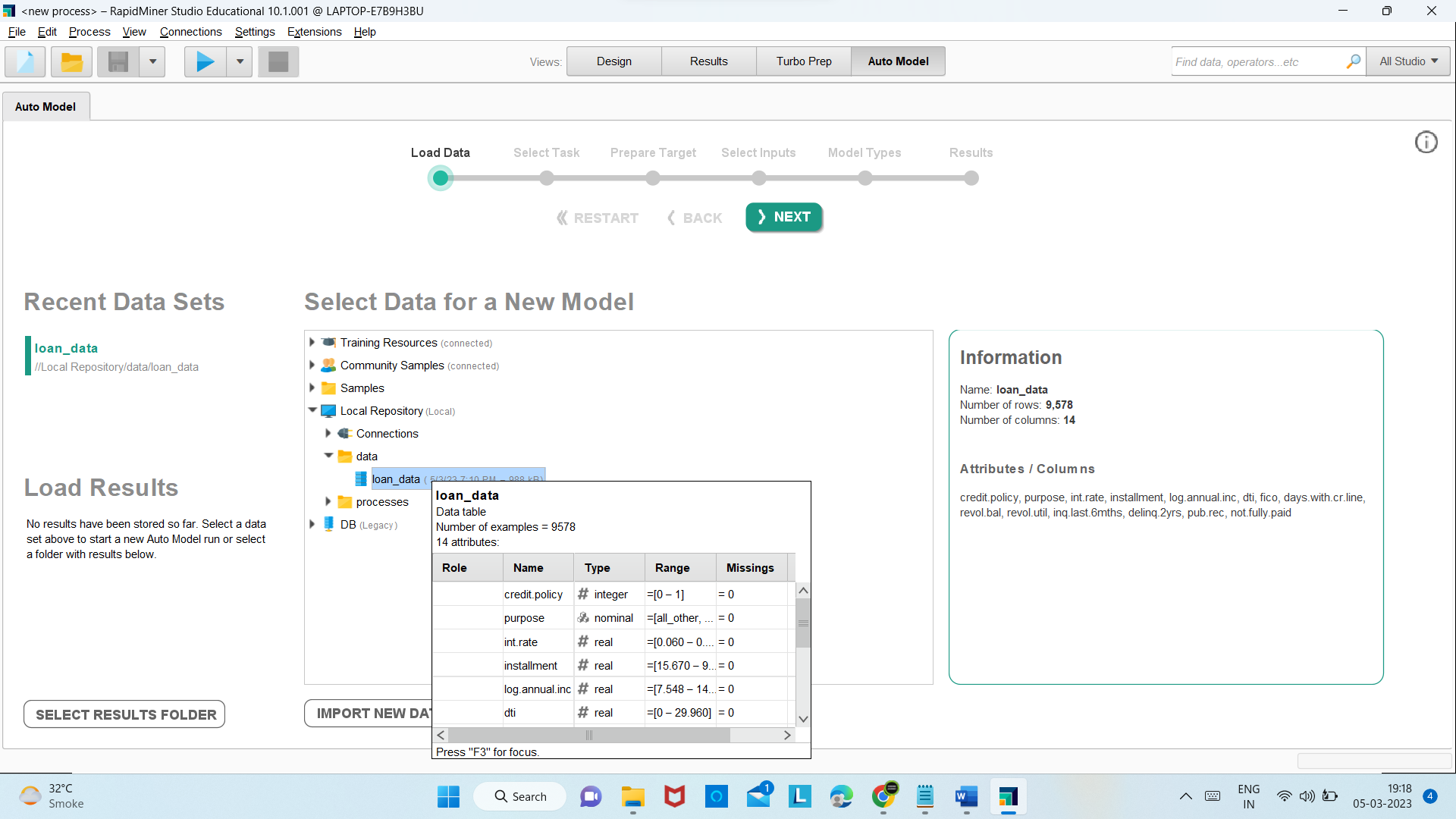
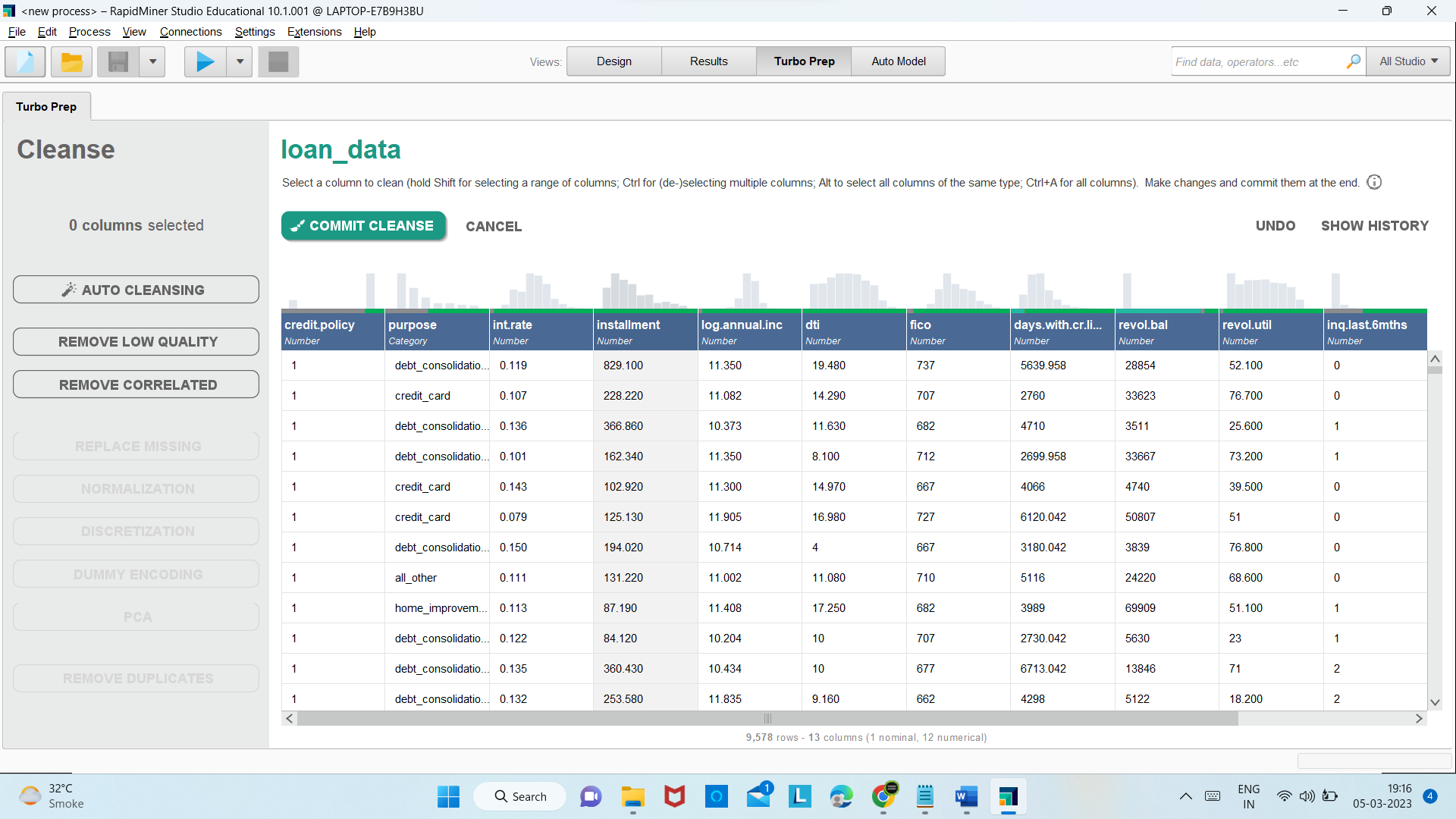
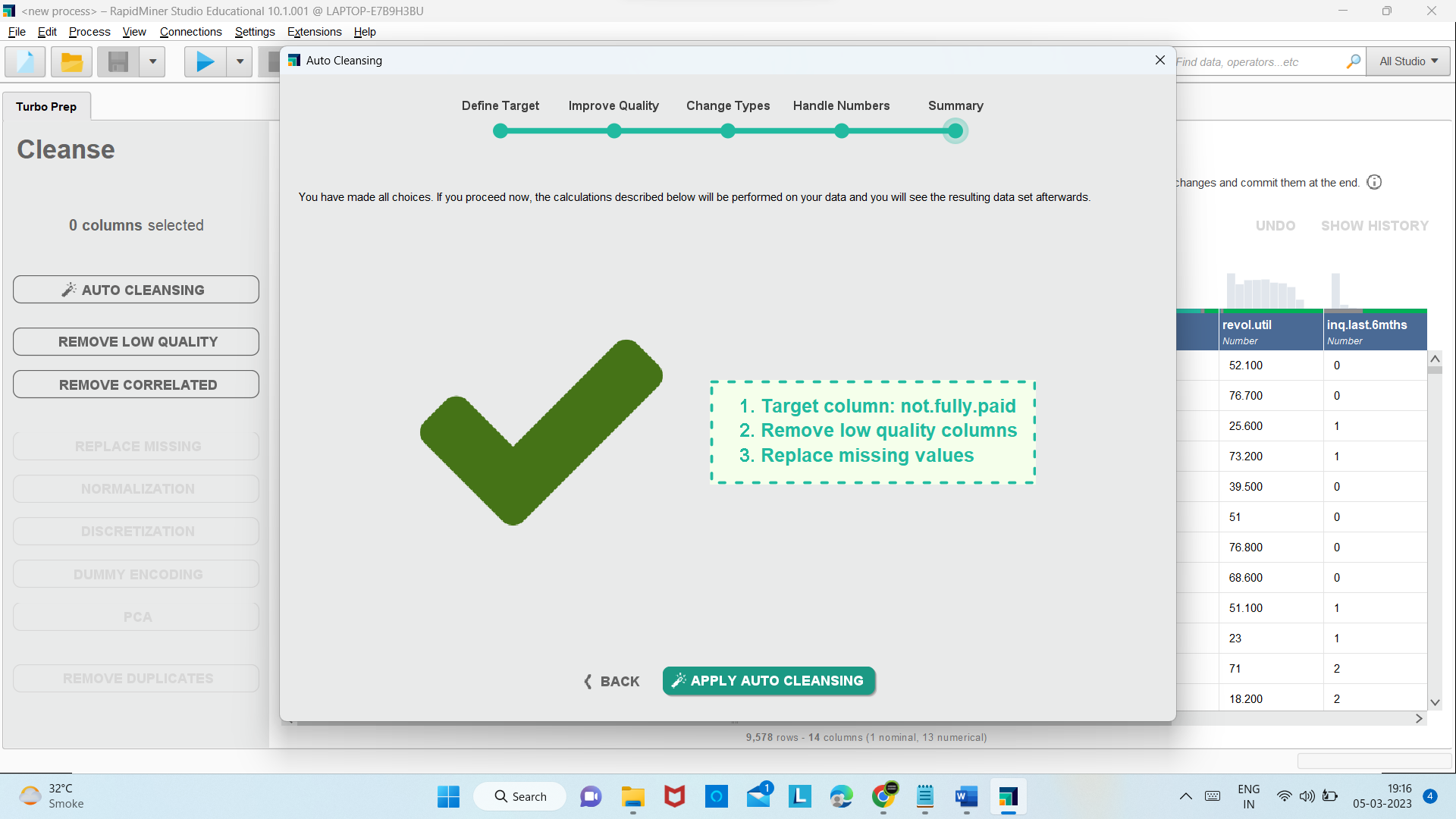
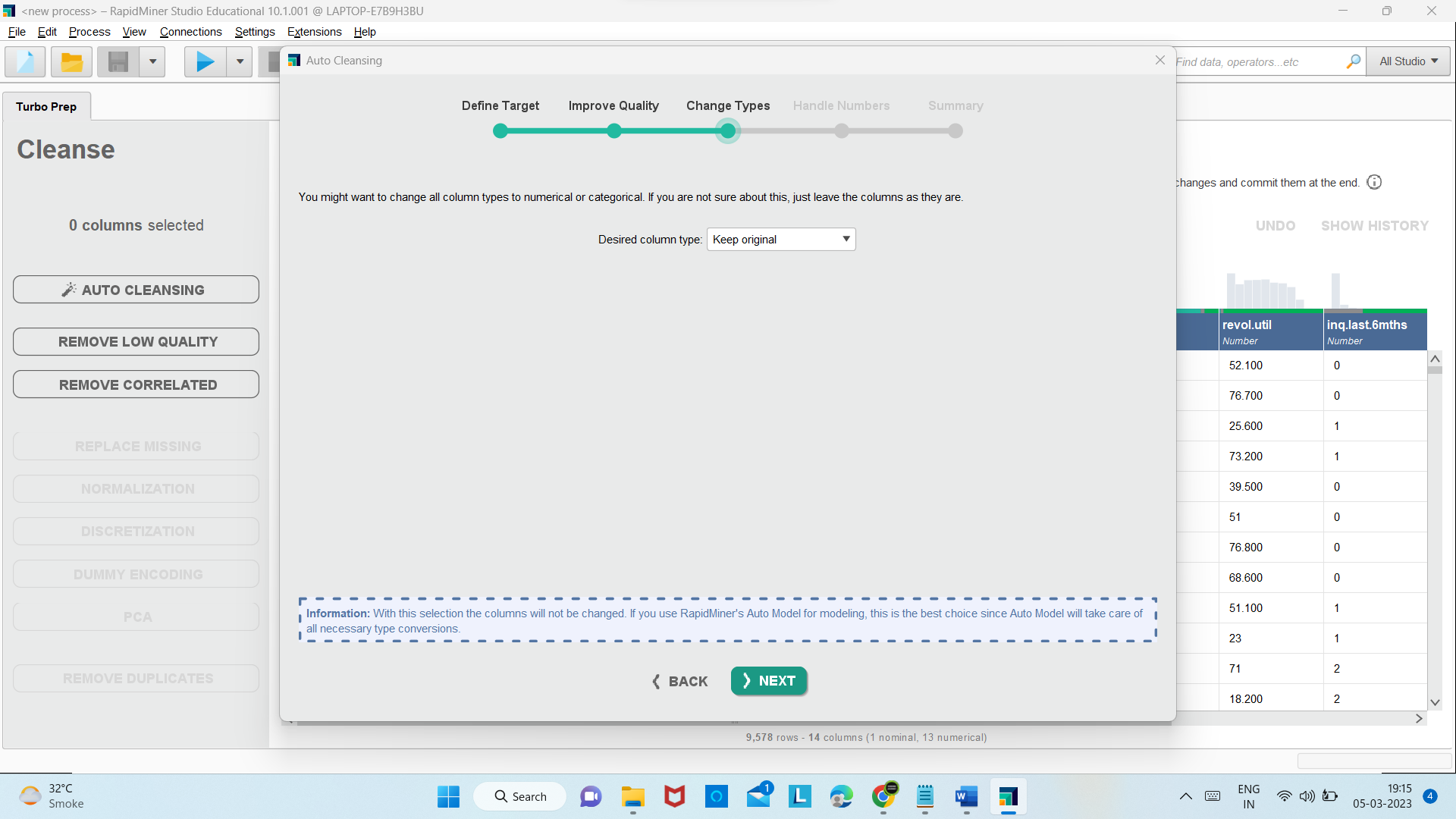
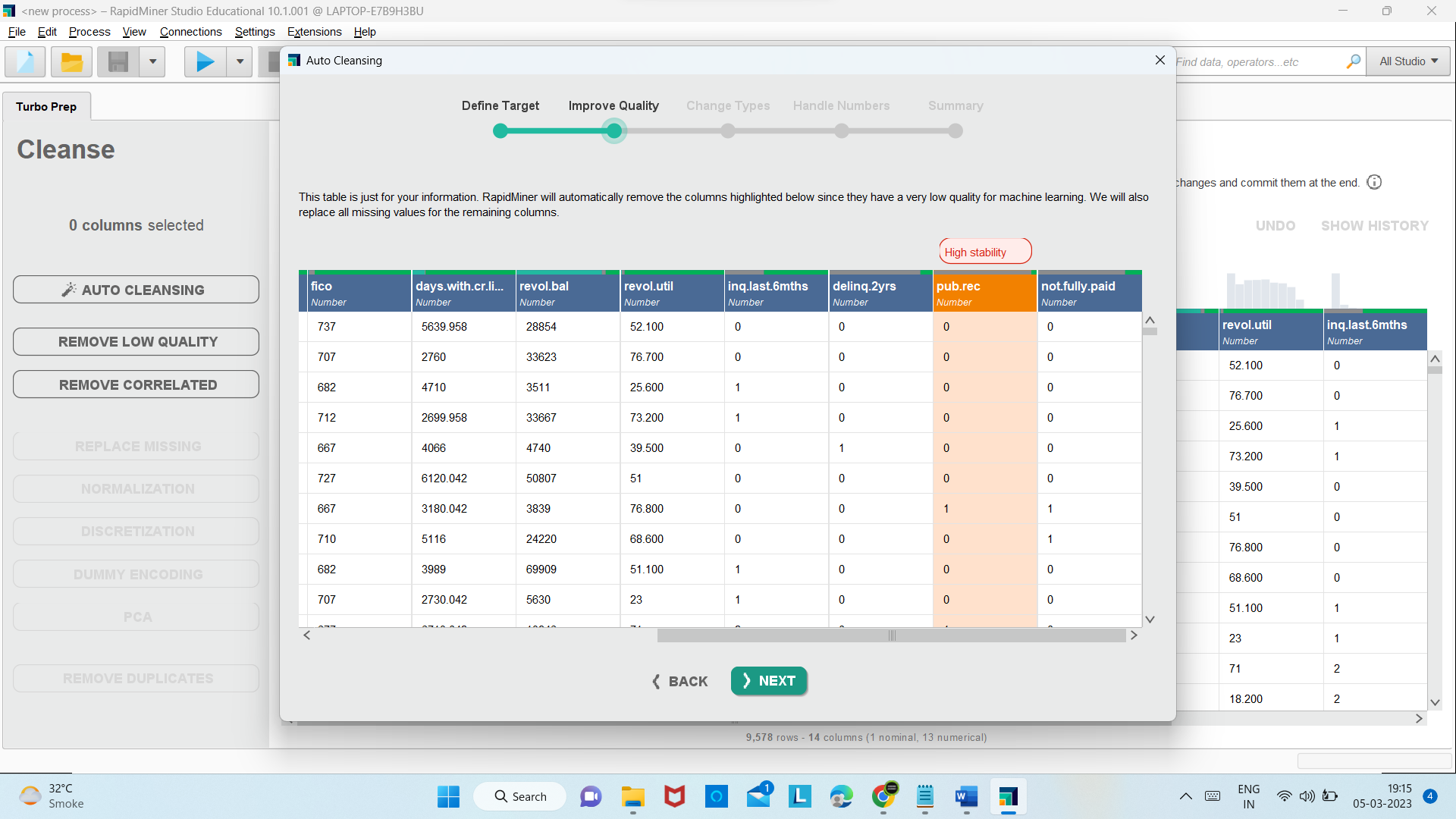
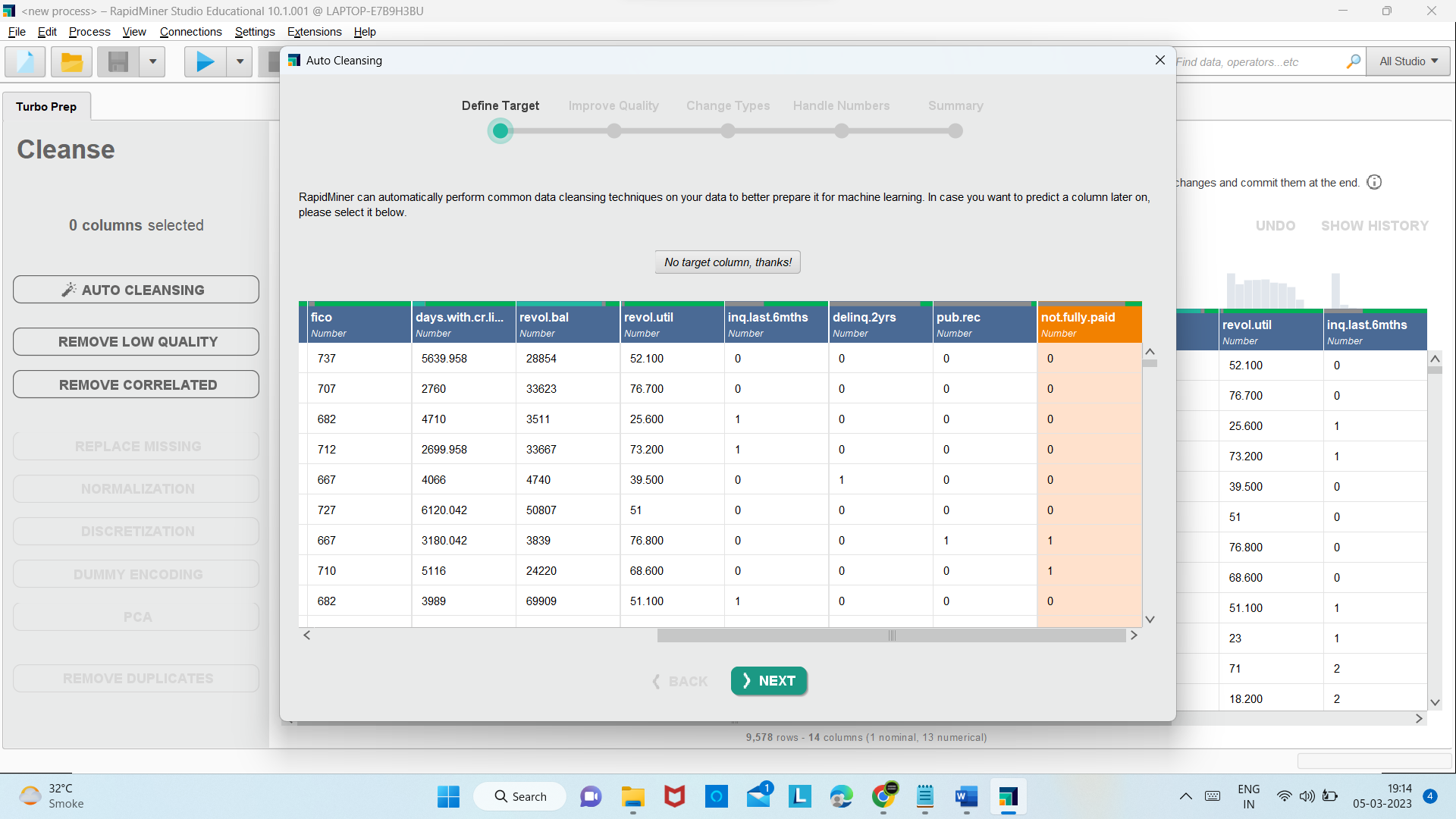
Naive Bayes: A probabilistic algorithm that makes predictions based on the Bayes theorem, which assumes that the presence of a particular feature is independent of the presence of other features.

K-Nearest Neighbors (KNN): A non-parametric algorithm that classifies data based on the majority class of its k-nearest neighbors in the feature space.

Classification algorithms have a wide range of applications, such as email spam filtering, image recognition, sentiment analysis, and medical diagnosis. Choosing the right classification algorithm depends on the specific problem and the nature of the data.

**Data Preprocessing in RapidMiner tool**:

* First create account on RapidMiner tool and then install software.
* Importing dataset from machine

Click on Auto Cleaning

**Conclusion**:

We successfully trained our loan\_data with classifier methods using Rapid Miner software.