BRIEF REPORT: MODEL TRAINING WITH K-FOLD VALIDATION

Dataset : Car Insurance Data

https://www.kaggle.com/datasets/sagnik1511/car-insurance-data

Dataset :

It's an algorithm that we use to train our machine to give a result. It is further divided in two subsets a Training dataset and testing data set. Training data requires some human(Judgement of people) involvement to analyze the data for machine learning models.

In the following assignment we opt such dataset which involves supervised leaning model. Training dataset is given some label to enrich the learning experience for the user/machine. Labeled dataset show the target, which is the outcome we want from our machine learning model.

Motivation :

Rapid minor is a tool which is used for Data Mining, Forecasting and Data loading for our machine learning model. For the following dataset we predicted the outcomes in very simple manner despite being complicated dataset or how big the dataset is, we able to make confusion matrix, recall data, accuracy and precision.
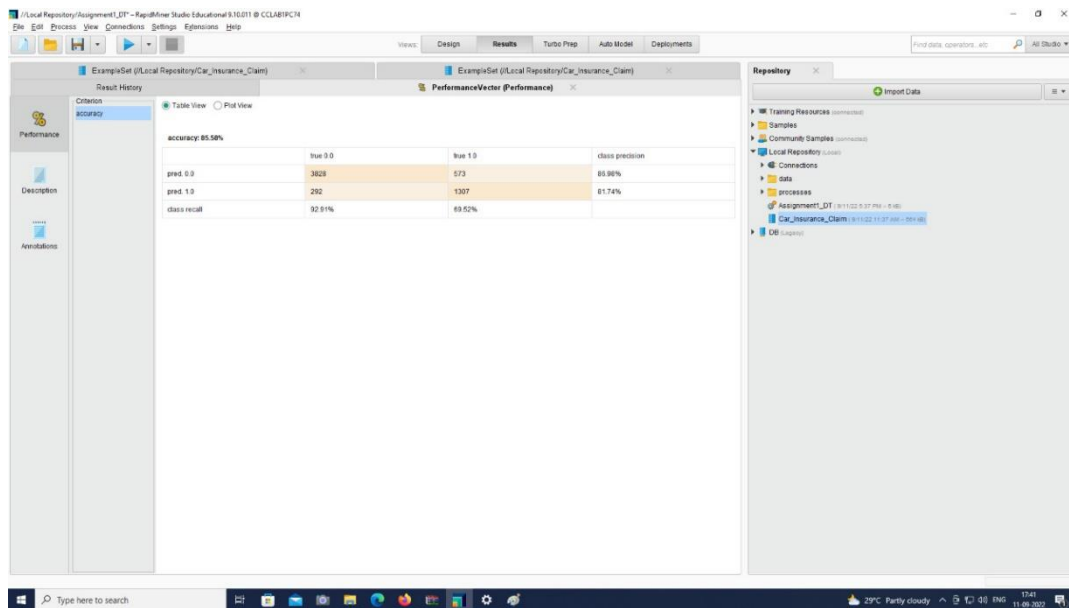
Features :

The dataset we choose for analysis contains various features and attributes like vehicle ownership, Vehicle year, Driving experience, Credit score. We choose label as an outcome. Here outcome 0 means a car have an accident and the owner does not take car insurance similarly 1 means a car have an accident and the owner does have a car insurance.

Tasks Performed :

First of all we go through various datasets which is mentioned on KAGGLE website. From the given datasets in binary classification we choose the Car Insurance Data. In which we have changed several feature's role to the binary which was Polynomial, Integer, Real, Float earlier. Labeled the outcome using rapid minor conversion.

Then training and testing are done as shown below.

- **60% Training – 40% Testing**

accuracy: 85.58%

	true 0.0	true 1.0	class precision
pred. 0.0	3828	573	86.98%
pred. 1.0	292	1307	81.74%
class recall	92.91%	69.52%	

- **70% Training - 30% Testing**

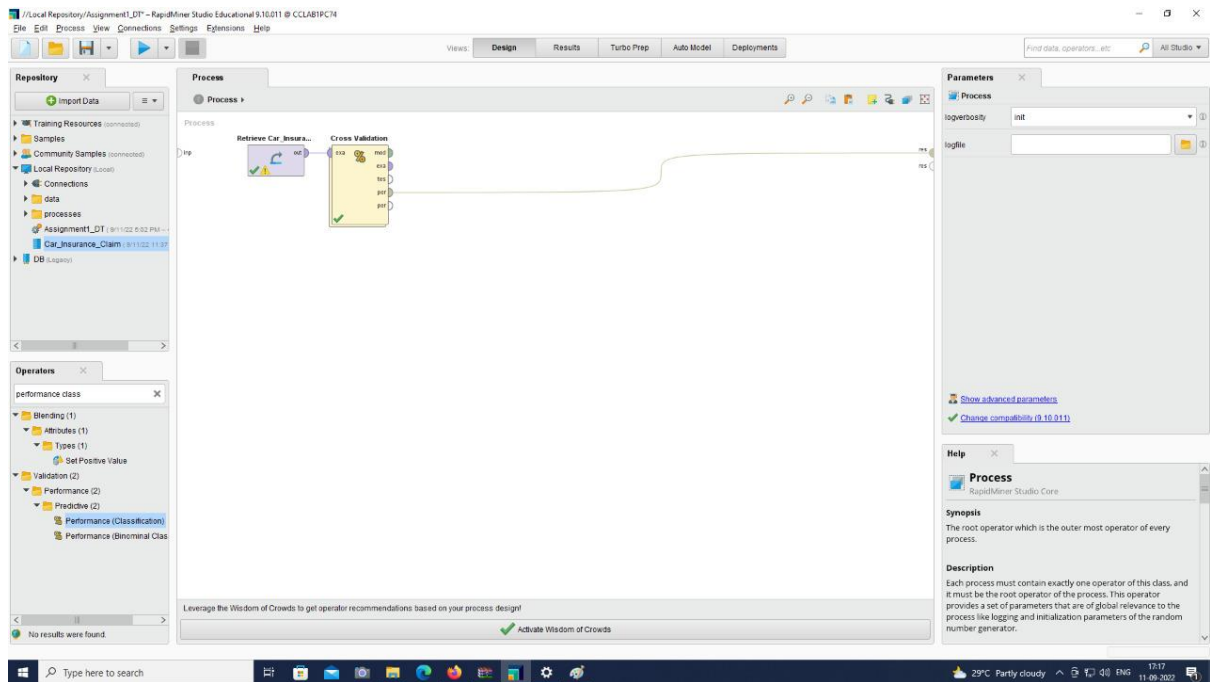- **80% Training – 20% Testing**

- **90% Training – 10% Testing**
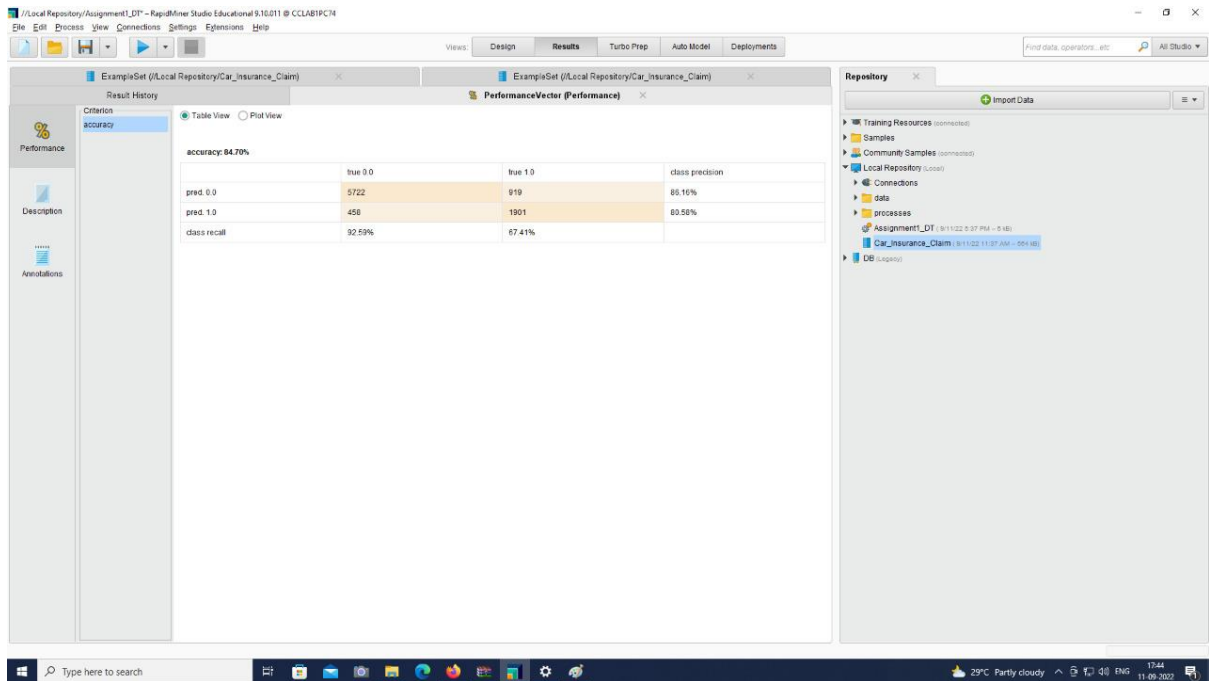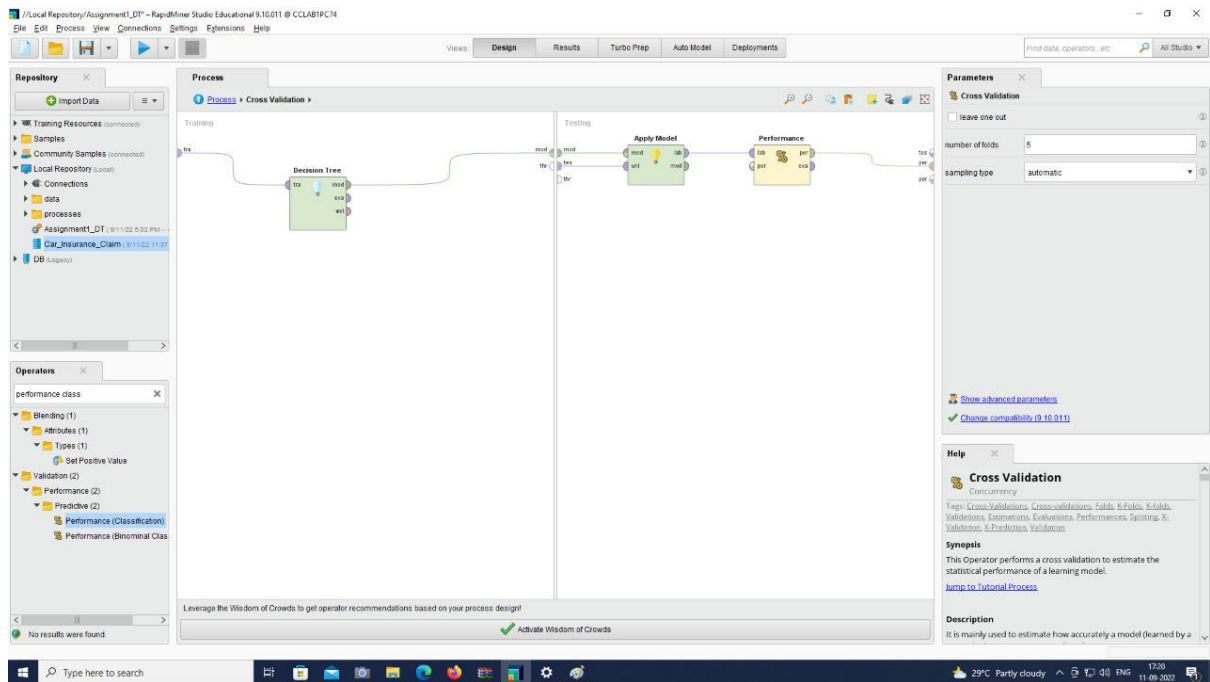

- **Split**
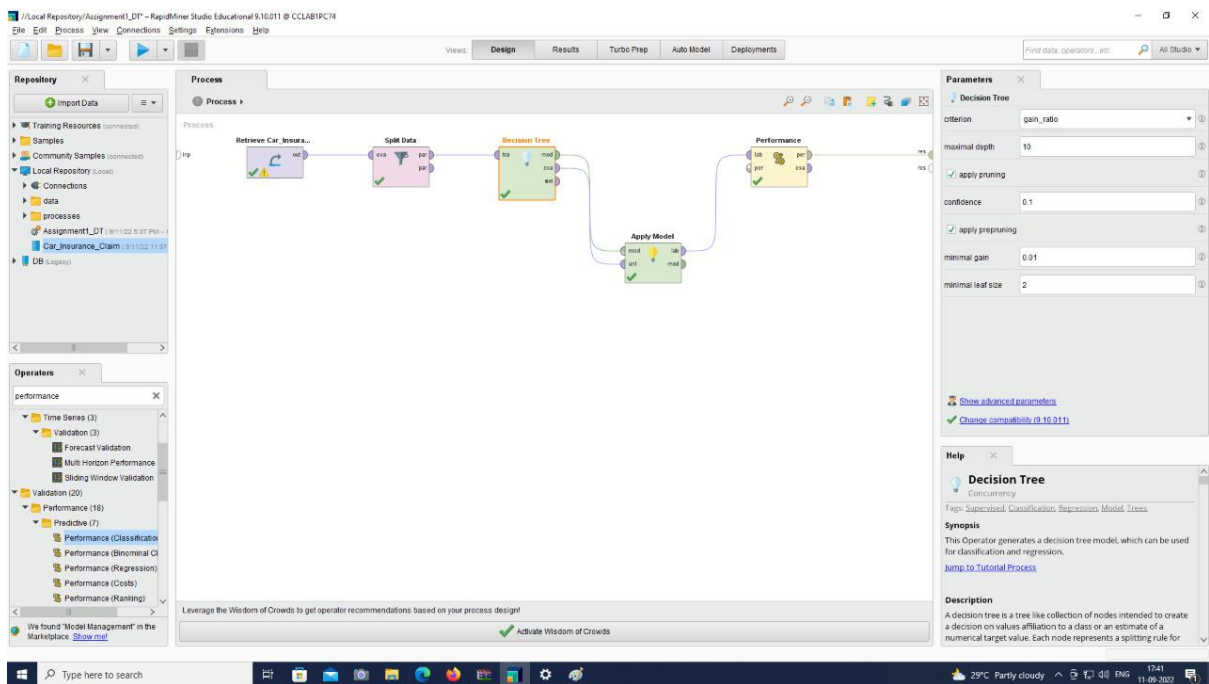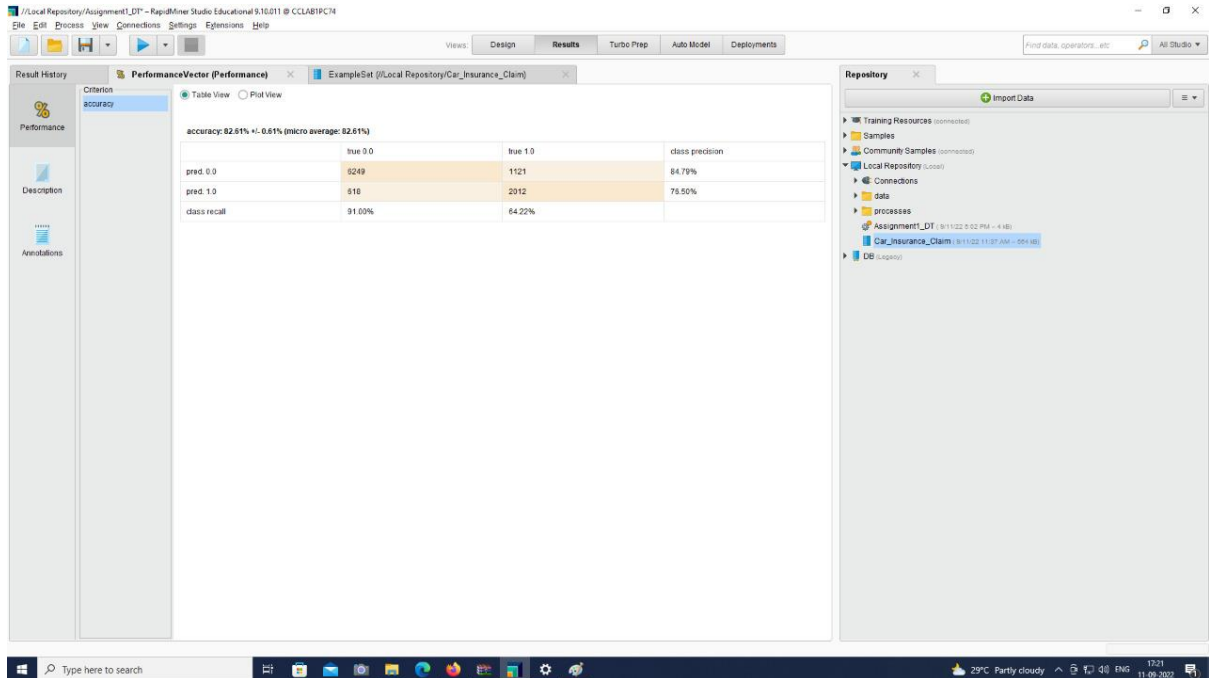

- **K- Fold Validation :**
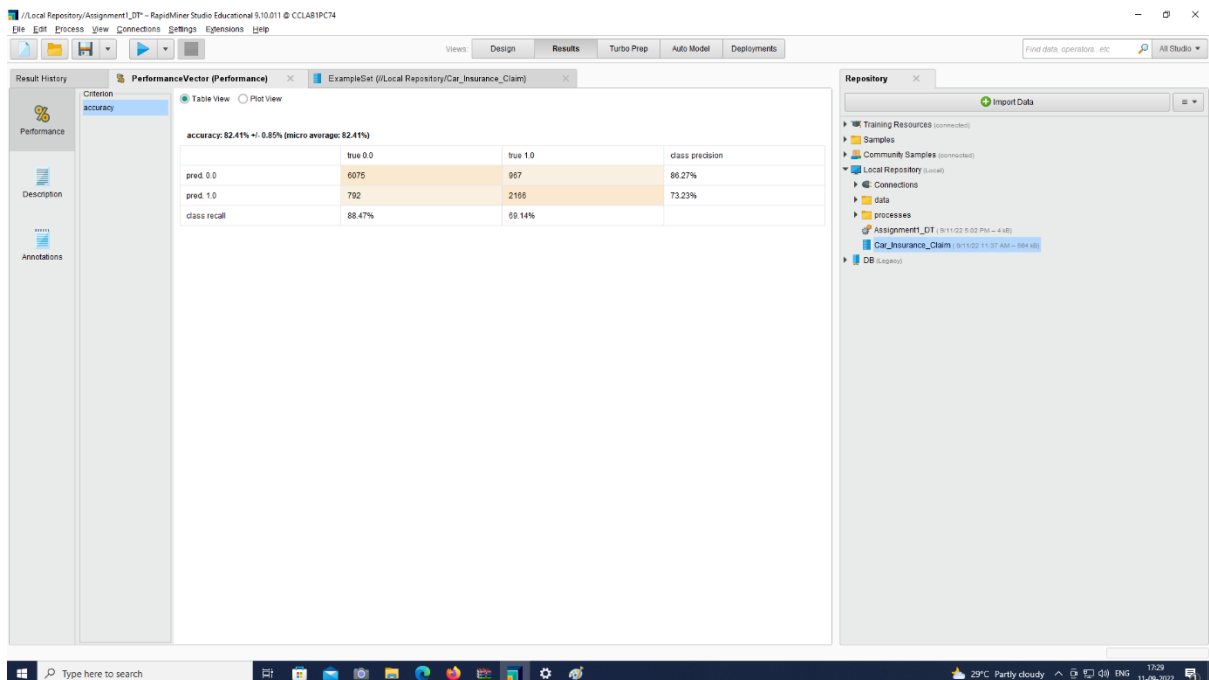
- **Inside cross validation :**

- **Result of K- Fold validation :**
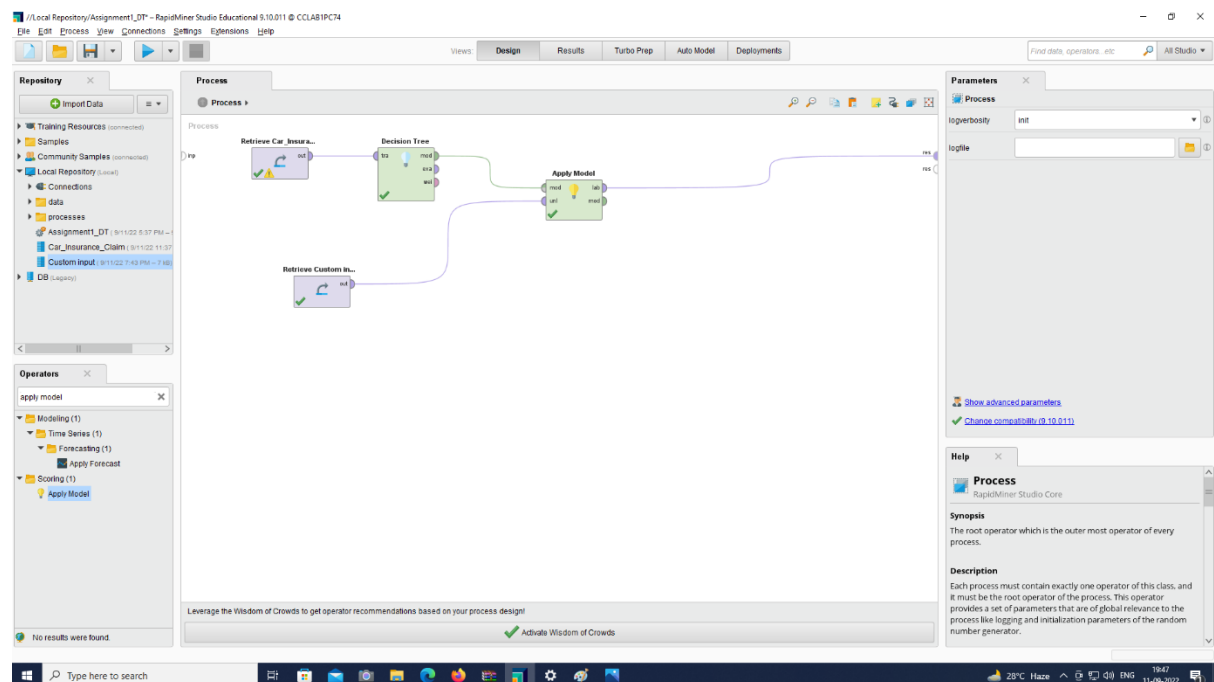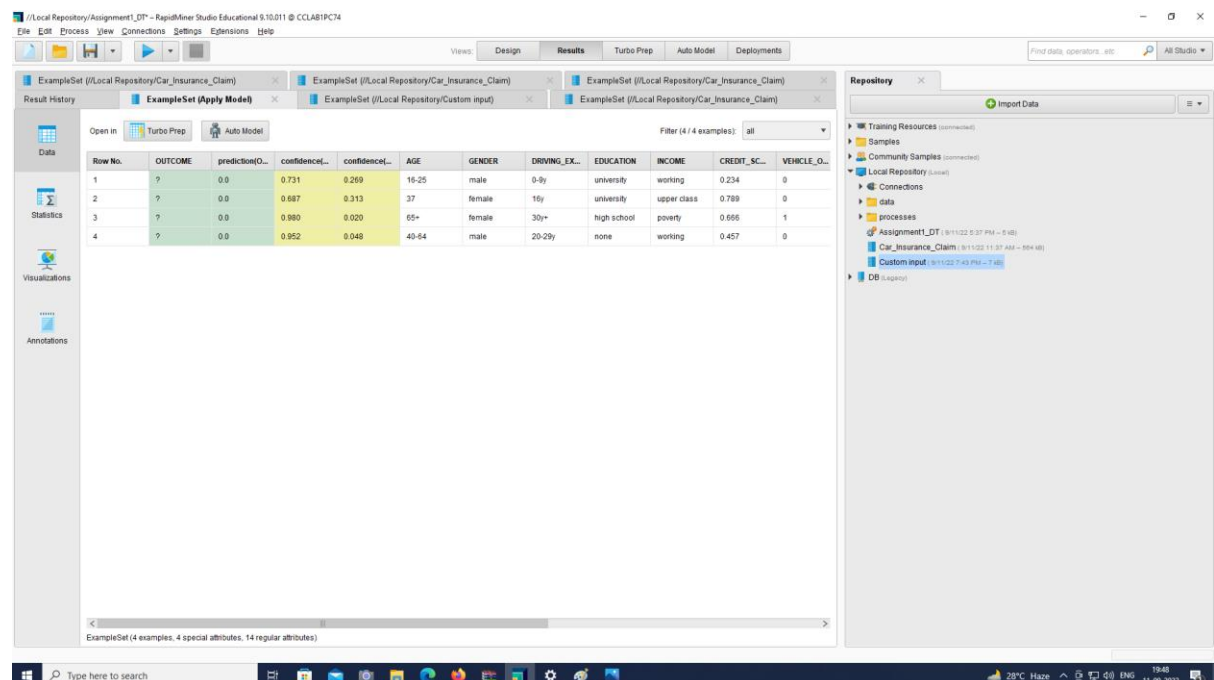
For k =5

For k =10,

Giving custom Inputs:

We have provided four custom inputs in order to test how well trained our model is,

Custom Results;

The Results of those 4 inputs have been shown here,

- **Research Paper Analysis :**

    Decision tree algorithm is one of the well known algorithm for data classification that has numerical and categorical attributes. The accuracy and performance of the leaned machine is highly dependent on the existing data and the method applied. The basic

concept of the decision tree algorithm is to transform data into decision tree using several rules. Data is expressed in form of attributes and features. The advantage of rapid minor software is that it gives flexibility to apply various algorithms in data features and attributes. Using Rapid Minor software we able to calculate the result of obtained accuracy about 91% and error of 9%.