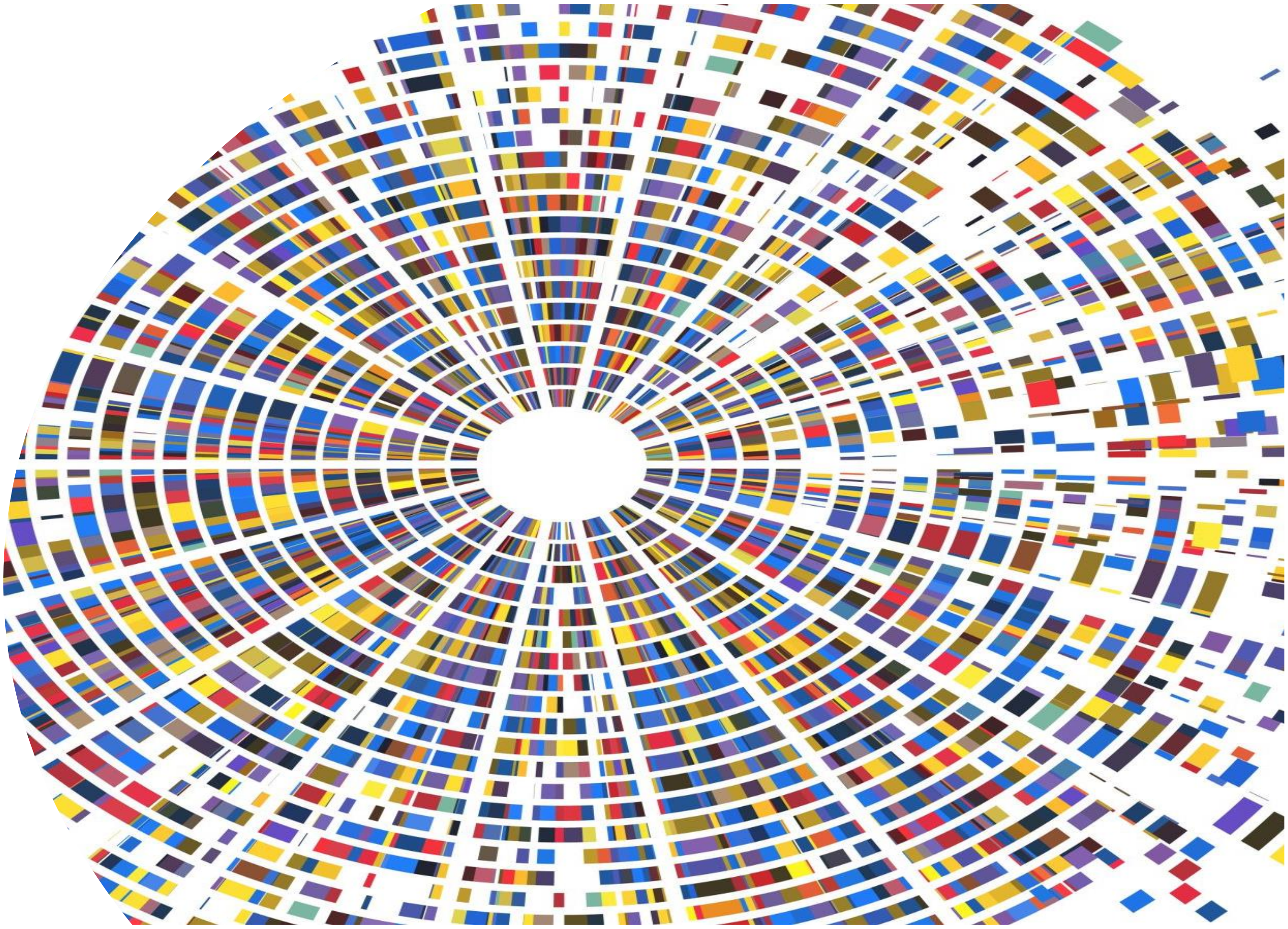
A Targeted Marketing Campaign: Data Analytics Problem



Sun Afolabi

Professor: Dr. Liang Chen

Comprehensive Exam Fall 2020

West Texas A & M University

November 18, 2020.

Table of Contents

Executive Summary

Goal and Motivation

It takes more effort, time, and money to win new customers than to retain the existing ones. This has been a universal accepted wisdom in business and marketing. Consequently, the onus has been on every business organization to device a means to cross-sell products to their existing customers in order to increase their share of wallet of their loyal customers. Failure to do this will naturally make the customers find an alternative business elsewhere which often than not leads to churn in favor of a company that offers diversified products and makes themselves a one-stop-shop. Success in doing this will make the organization keep its customers. In this project, we devise a marketing campaign at a targeted population of a bank's customers.

A target market is defined as "a group of customers within a business's serviceable available market at which a business aims its marketing efforts and resources. A target market is a subset of the total market for a product or service." (*Target Market*, 2020) The goal is to predict the likelihood that 20 out of the 200 targeted customers will likely accept a loan offer. The prediction will be established through data mining models and based on the given attributes.

Method and Data

This project is on the marketing analytics domain and to be able to achieve the objective, secondary data supplied by the professor is utilized. The two datasets are posted on the West Texas A&M University web class blackboard. On the original datasets, data visualization and descriptive analytics were conducted. This leads to the next step; the data was prepared by removing variables that could impact the outcome of the prediction. The predictive analytics component of the project was completed by using the training dataset with a predictive dataset to create a decision tree on RapidMiner while the manual identification of the outcome of the data analysis was done using Microsoft excel.

Key Findings

Using the decision tree, there are 93 people out of 131 that are predicted to accept the offer. This a probability of 0.71. The 20 people who will likely accept the offer come from the group that is not married and has no mortgage. And more importantly, they should be living in the Inner City to increase the expected return from this category of customers for the bank.

Conclusion

In conclusion, running a productive marketing campaign is the desire of every organization. Running a barren campaign is a major concern. In running a successful campaign, many unexpected factors can influence the decision of customers who have responded "Yes" to a previous campaign and have responded "Yes" to the current campaign. Such inherent factors are not the object of consideration in this analytics project. The output of this analysis can be utilized in any financial institution or marketing department of a company who desires to run a targeted campaign. The existence of historical behavior data of customers helps predict the future behaviors of customers' purchase habits.

Business Understanding

It is the desire of every organization to grow in profit every year. This ensures they continue in business as a going concern. This makes it indispensable for an organization to hold strategy sessions yearly, quarterly, and some monthly to formulate policy and implementation procedures that will lead to customer acquisition, retention, and profit maximization. To achieve this lofty aim, organizations research their markets and customer patronage that makes it possible for them to create market segmentation. Market segmentation is defined as "a process of dividing a heterogeneous market into relatively more homogenous segments based on certain parameters like geographic, demographic, psychographic, and behavioural" (Wikipedia contributors, 2020).

Because one size does not fit all, there is wisdom for market segmentation. It allows for stronger marketing messages, targeted digital advertising, development of effective strategies, better response rates, and lower acquisition costs, attracting the right customers, increasing brand loyalty, branding differentiation from other competitors, identifying niche markets, staying on message, driving growth, enhanced profits, and ideal product development. In fact, according to a study by Bain & Company, "81% of executives found that segmentation was crucial for growing profits. Bain also found that organizations with great market segmentation strategies enjoyed a 10% higher profit than companies whose segmentation wasn't as effective over a 5-year period" (Market Segmentation: Everything to Know in 2020 //, 2020). All these points to a fact; market segmentation is good and effective for a business to be successful.

Problem to Solve

Kaplan and Norton (1992) argue "Managers must identify the customer and market segments in which the company competes and clarify the appropriate measures of performance in these targeted segments. Outcome measures, such as customer satisfaction, customer retention, new customer acquisition, customer profitability, and market share, must be linked to the targeted customer segments in which the business anticipates its greatest potential for growth and profitability"

This analytics project is premised on bank profitability through the effective use of market segmentation. This will allow the bank to increase its market share and the share of the wallet of the customers. The customers have been identified. The propensity of each customer to respond positively to the marketing campaign is the basis of this project. While customer loyalty plays a decisive role in customer identification, even among loyal customers, there is the need to know and understand those who might be interested in responding to a specific loan the bank is offering. This is the basis of the target campaign and the whole essence of this project and therefore the problem it seeks to solve.

Motivation

According to an outbound agency that provides statistics on organizations and customer acquisition, acquiring a new customer costs anywhere between five times to twenty-five times more than retaining an existing customer. More so, increasing customer retention by 5% can increase profits from 25-95% while the "success rate of selling to a customer you already have is

60-70%, while the success rate of selling to a new customer is 5-20%” (Landis, 2020). It makes logical sense for a smart organization to focus more energy on customer retention by cross-selling products that can increase its share of the customers’ wallet than devoting all energy and attention to acquiring the new ones. The agency goes further to say: “U.S. companies lose $136.8 billion per year due to avoidable consumer switching” (Landis, 2020). Needless to say, the need to keep the existing customers from escaping cannot be overemphasized. Importantly, this project is a requirement for the final comprehensive examination required for the graduate school of Masters in Computer Information Systems and Business Analytics at West Texas A&M University.

Opportunity/Challenge

The goal of this project is to predict 20 customers who will respond positively to the marketing campaign the bank is embarking on using the existing data based on the customers' historical behaviors. By analyzing the datasets that contain hundreds of records through data analysis in tools such as Excel and RapidMiner, it is possible to unveil the features of the customer in the dataset, predicting the 20 individuals with a high probability of saying “Yes” thereby accepting the loan for the success of the marketing campaign.

Brief Action Plan

To be able to fulfill the goal of this project, the given dataset will be subjected to the data analytical tools: Microsoft excel, RapidMiner. Once the data is prepared, the training dataset will be used to create a decision tree. This will help to determine the best performing model on the prediction data set to predict the likelihood that a given individual in the prediction dataset will accept the loan.

I. Research Method/ Design

The project method is referred to as a “classic machine learning procedure” that utilizes the scientific paradigm of induction and deduction (University of Eastern Finland, n.d.). In the inductive phase, machine learning models on secondary datasets retrieved from WTAMU blackboard are trained.

The inductive phase helps create models that learn general rules from our dataset; general rules which accept “positive examples” and reject “negative examples” via a process called supervised machine learning. The structure of the dataset makes the decision tree the ideal models

In the deductive phase (predictive phase), the best performing model from the inductive phase is used to make predictions using new data (second dataset). This schema includes data collection, data preparation, model building, model evaluation, and model deployment to make predictions as depicted in Figure 1.

The data analysis was conducted using descriptive analytics techniques. This gives room for data visualization for a graphical understanding of different attributes available in the dataset. The target variable is a response to know those who answered “yes” to the previous marketing campaign and those who answered “no” to them. The predictors or independent variables are age, sex, region, income, married, save_act, and current_act. car and mortgage are controlled variables. The supervised machine learning method enables the understanding and identification of the attributes that are important predictors of the target variable “response.”

This uses relevant attributes to predict the probability that a customer will accept the loan. The project is marketing analytics. Descriptive analytics (D) is used to visualize and better understand the data while predictive analytics (P) is used to make a prediction. The outcome will bring Innovation (I) and Agility (A) to bank loan marketers, which will increase their Productivity (P). By developing a predictive model, bankers and other financial service organizations can expeditiously and accurately predict which customers will accept a loan offering with alacrity. This will allow them to quickly make such loans available to the individuals who will not refuse. It eliminates wasting spending on campaigns that would have been directed to unwilling customers who will ultimately reject such loans.

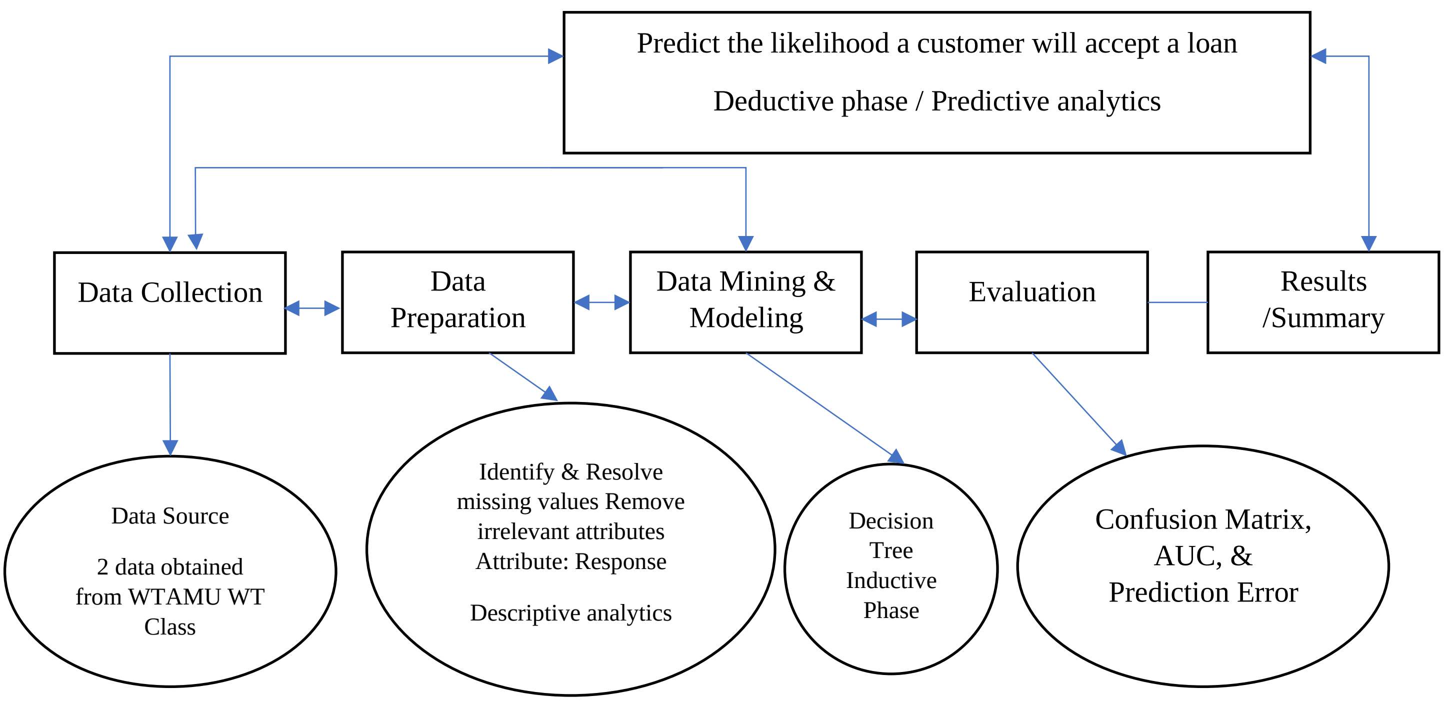


Figure 1: Predictive analytics schema

Variable	Definition	Measurement
Id	To identify each customer	Unique ID
Age	A customer's age (years old) when receiving the loan offer	Number in Years
Sex	A customer's gender	Male / Female / Other
Region	The region in which a customer is living when receiving the loan offer	Inner City, Rural, Suburban, Town
Income	A customer's annual taxable income when receiving the loan offer	Amount in numbers
Married	Is this customer married when receiving the loan offer?	Yes / No
Children	The number of children that a customer has when receiving the loan offer	Count in numbers
Car	Does this customer own a car when receiving the loan offer?	Yes / No
Save_act	Does this customer have a saving account when receiving the loan offer?	Yes / No
Current_act	Does this customer have a current loan account with the bank?	Yes / No
Mortgage	Does this customer have a mortgage when receiving the loan offer?	Yes / No
Response	Does this customer respond to the loan offer	Yes / No

Table 1: Variable definition and measurement

II. Data Description

Among the two datasets, all attributes include "ID", and 8 independent variables: age, sex, region, income, married, children, save_act, and current_act. car and mortgage are controlled variables. Response is the target variable present in the training dataset but empty in the predictive dataset. The training dataset contains 601 records; the prediction dataset contains 200 records.

Attribute Name	Description	Data Type
Id	To identify each customer	Numerical
Age	A customer's age (years old) when receiving the loan offer	Discrete/ Integer
Sex	A customer's gender	Categorical/Binomial
Region	The region in which a customer is living when receiving the loan offer	Categorical/Binomial
Income	A customer's annual taxable income when receiving the loan offer	Ordinal/Continuous
Married	Is this customer married when receiving the loan offer?	Categorical/Binomial
Children	The number of children that a customer has when receiving the loan offer	Discrete/integer
Car	Does this customer own a car when receiving the loan offer?	Categorical/Binomial
Save_act	Does this customer have a saving account when receiving the loan offer?	Categorical/Binomial

Current_act	Does this customer have a current loan account with the bank?	Categorical/Binomial
Mortgage	Does this customer have a mortgage when receiving the loan offer?	Categorical/Binomial
Response	Does this customer respond to the loan offer	Categorical/Binomial

Table 2: Training dataset attributes

III. Data Preparation

Data Preprocesses and Original Data Set Descriptive Analysis

Data preparation takes time. It takes 80% of the time for data scientists when working on a project (Press 2016). This is expected as the preparation must be right in order to have reliable data analysis. The data was cleaned to allow the correct outcome.

The original train dataset contains missing values for Age, Income, Car variables: Age has missing values of 8 which represents 0.12% of the train dataset records. Car missing values are 5 that represents 0.07% of the train data set records. The Income has the largest missing values of 412 that represents 6.23% of the train data set records. In all 425 missing values is equivalent to 6.43% train dataset records. The “children” attribute is present in the training dataset but absent in the predictive dataset. To balance the data, this attribute was deleted. The implication of this is the number of children a customer has will not determine his response on the decision tree on RapidMiner. The “Response” attribute in the predictive dataset has empty values.

IV. Results and Findings

Descriptive Statistics

Descriptive Statistics was carried out on the original data set. A summary of the descriptive statistics for both datasets is below.

The Income mode on the training dataset is $17, 546. There is no income mode on the predictive dataset. The 412 missing values on income attribute in the training dataset account for the difference visible on the Income train and predictive dataset.

Table 3: Numerical variables descriptive statistics

	Age Train	Age Predict	Income Train	Income Predict	Children Train
Mean	46.82	41.77	26937.71	26766.99	1.17
Standard Error	2.92	1.02	928.30	875.73	0.17
Median	41	41.50	24477.50	23908.35	1
Mode	43	60	17546	#N/A	0
Standard Deviation	71.16	14.38	12762.07	12384.70	4.13

Sample Variance	5063.61	206.83	162870446.99	153380757.56	17.09
Kurtosis	162.41	-1.14	-0.07	-0.16	524.78
Skewness	12.43	0.07	0.80	0.71	22.16
Range	981	49	53141.25	55516.40	99
Minimum	18	18	7606.25	5960.40	0
Maximum	999	67	60747.50	61476.80	99
Sum	27762	8354	5091227.57	5353397.05	706
Count	593	200	189	200	601
Confidence Level(95.0%)	5.739047375	2.005351131	1831.230753	1726.902354	0.331186567

Decision Tree Results

The importance of the decision tree is fully experienced in this project. This is because when all means of solving the question came to halt. It was the decision tree that assisted to remove the impasse and get a meaningful solution. No wonder, it is said: “Decision trees provide an effective method of Decision Making because they: Clearly, layout the problem so that all options can be challenged. Allow us to analyze fully the possible consequences of a decision. Provide a framework to quantify the values of outcomes and the probabilities of achieving them” (Decision Tree Analysis: Choosing by Projecting “Expected Outcomes,” 2020).

This project would have been impossible without the use of a decision tree on RapidMiner. The result of the decision tree is below. The result reveals 93 people out of 131 are predicted to accept the offer. This is a probability of 0.71. The ideal person is not married and has no mortgage. Hence, 20 customers come from this group. Most importantly, they should be living in the Inner City for the maximum expected return.

Decision Tree Model

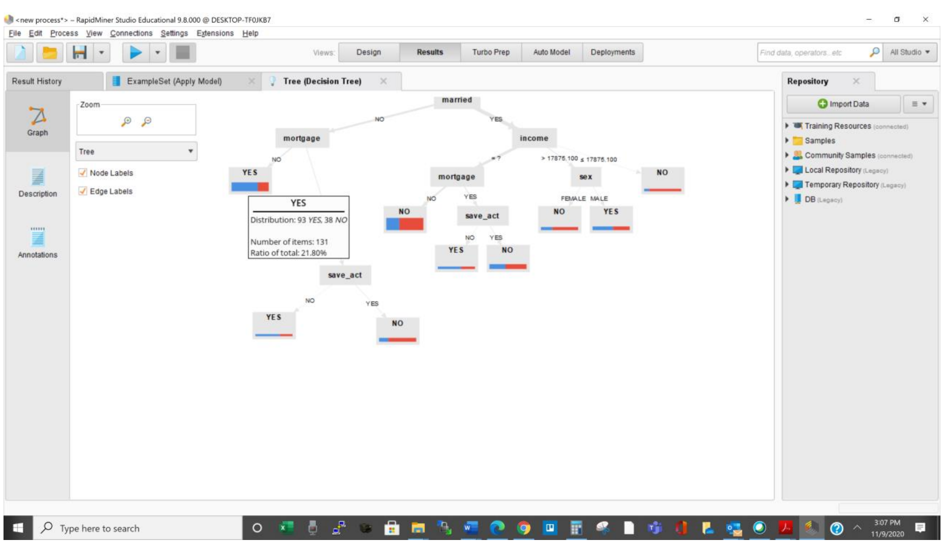


Figure 2: Decision Tree Model

Figure 2 depicts the tree of the decision tree model. We can see that married is the root node, but it also appears to be a split node at some other parts of the tree. According to the decision tree, for customers who are not married, have no mortgage, 93 respond yes, 38 no. Customers that are not married, have mortgage no savings account, 15 respond yes and 8 no. Customers that are not married, have a savings account, 13 respond yes and 38 no. Conversely, customers that are married, have no mortgage 64 respond yes, 118 no. For customers that are married and have a mortgage but have no savings account, 18 respond yes, 11 no. Customers that are married and have a mortgage but have a savings account, 29 respond yes, 38 no. Customers that are married, have income greater than $17,875.100 who are female 12 respond yes, 27 no. Customers that are married, have income greater than $17,875 who are male 27 respond yes, 22 no. Customers that are married, have income less than $17,875 4 respond yes, 26 no. The degree of yes response to the bank campaign is obviously a function of individual customers' income.

Prediction Results

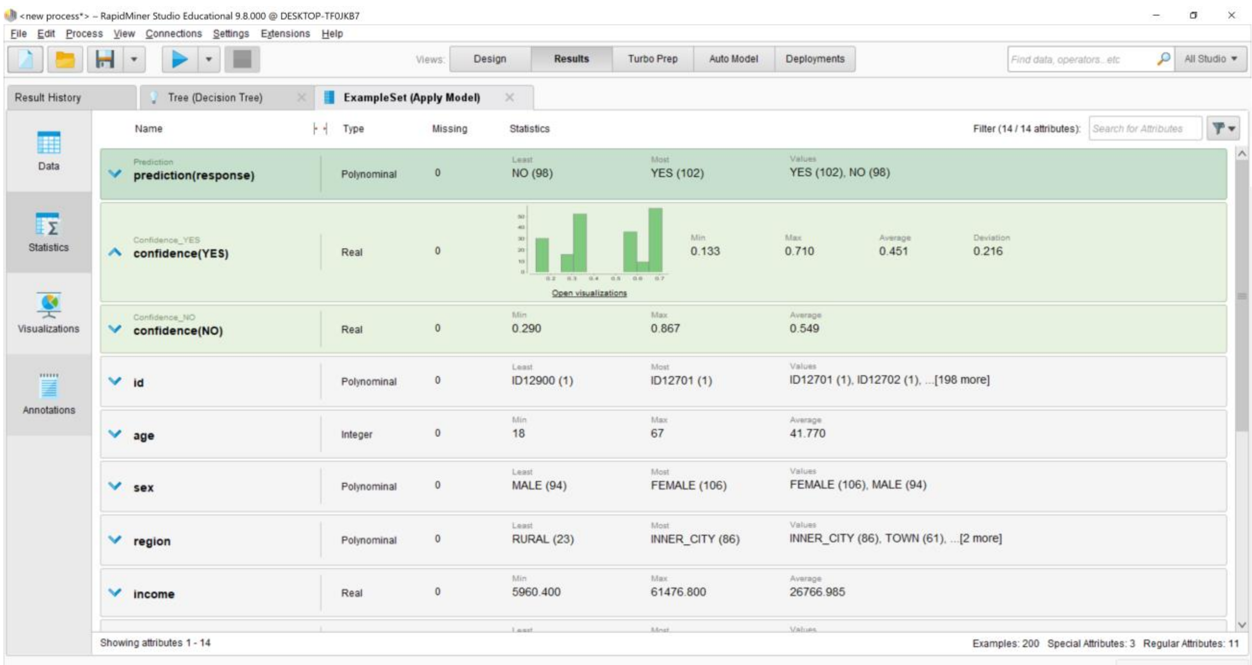


Figure 3: Decision Tree (Apply model prediction example set)

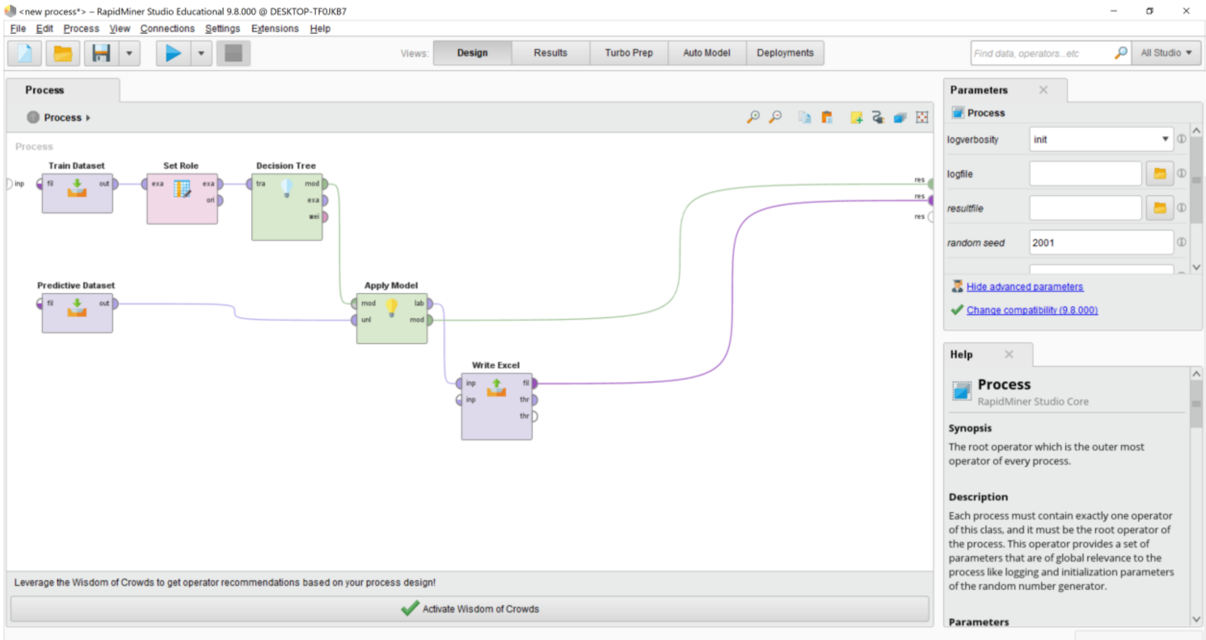


Figure 3. Export to Excel from RapidMiner

Findings

After the processed data are exported from RapidMiner to excel, the 20 required customers are sorted manually. The list of the 20 customers is listed below. In all, 25 customers are qualified for the 0.71 probability but 5 were expunged based on the paucity of income. This leaves us with the 20 lists on the table below. The 0.71 probability is 71%. The customers live in Inner City. As a result, the expected return from each customer is $10,000 * 71% => $7,100. For the 20 customers: $7,100 * 20 => $142,000.

These findings support the argument "companies that take a proactive stance in using customers and suppliers as a key source of inspiration, rather than merely monitoring and imitating what competitors are doing, are those that can gain greater rewards in the marketplace and earn a higher market share with better brand awareness in their respective industries" (Dawson & Andriopoulos, 2017, p. 774).

s/n	id	age	sex	region	income	married	car	save_act	current_act	mortgage	confidence (YES)	prediction (response)
1	ID12897	19.0	MALE	INNER_CITY	17,906.80	NO	YES	YES	NO	NO	.7	YES
2	ID12771	23.0	MALE	INNER_CITY	18,233.10	NO	YES	NO	NO	NO	.7	YES
3	ID12773	26.0	FEMALE	INNER_CITY	23,925.60	NO	YES	NO	YES	NO	.7	YES
4	ID12770	28.0	MALE	INNER_CITY	17,846.20	NO	NO	YES	YES	NO	.7	YES
5	ID12832	28.0	MALE	INNER_CITY	23,505.80	NO	NO	NO	YES	NO	.7	YES
6	ID12900	34.0	MALE	INNER_CITY	25,843.10	NO	NO	YES	YES	NO	.7	YES
7	ID12757	38.0	MALE	INNER_CITY	30,943.70	NO	YES	YES	YES	NO	.7	YES
8	ID12753	40.0	FEMALE	INNER_CITY	32,089.90	NO	NO	YES	YES	NO	.7	YES
9	ID12709	42.0	MALE	INNER_CITY	33,584.90	NO	YES	YES	NO	NO	.7	YES
10	ID12766	44.0	FEMALE	INNER_CITY	31,833.70	NO	NO	YES	YES	NO	.7	YES
11	ID12830	46.0	FEMALE	INNER_CITY	17,049.30	NO	NO	YES	NO	NO	.7	YES
12	ID12817	47.0	MALE	INNER_CITY	36,766.00	NO	NO	NO	YES	NO	.7	YES
13	ID12759	54.0	FEMALE	INNER_CITY	27,775.70	NO	NO	NO	YES	NO	.7	YES
14	ID12831	56.0	FEMALE	INNER_CITY	41,523.00	NO	NO	YES	YES	NO	.7	YES
15	ID12779	57.0	MALE	INNER_CITY	39,576.30	NO	YES	YES	YES	NO	.7	YES
16	ID12814	57.0	MALE	INNER_CITY	37,548.00	NO	NO	NO	YES	NO	.7	YES
17	ID12804	60.0	MALE	INNER_CITY	19,849.00	NO	NO	YES	YES	NO	.7	YES
18	ID12746	61.0	MALE	INNER_CITY	26,527.90	NO	NO	NO	YES	NO	.7	YES
19	ID12747	65.0	FEMALE	INNER_CITY	58,002.20	NO	NO	YES	YES	NO	.7	YES
20	ID12811	67.0	MALE	INNER_CITY	39,137.00	NO	YES	YES	NO	NO	.7	YES

Table 4: The List of 20 Customers

V. Discussion

“Predictive analytics are used to determine customer responses or purchases, as well as promote cross-sell opportunities. Predictive models help businesses attract, retain, and grow their most profitable customers. Improving operations. Many companies use predictive models to forecast inventory and manage resources” (Predictive Analytics: What It Is and Why It Matters, 2020). The task of identifying the 20 customers out of the morass of 8,611 is daunting and intensive. The use of predictive analytics help simplifies the process and facilitates the eventual arrival at the expected solution. From the findings, 25 customers were identified as having a high probability of 0.71. The decision to drop 5 based on either salary or age was given a deep thought. The lower salary is favored because of the propensity to repay the loan. Customers with a higher age of above 60 have a higher income. Hence, they are included in the final result of the 20 customers.

VI. Limitations/Areas for Future Improvement

The data analytics project is limited for two reasons. There are 415 missing values in the train datasets. The attribute “children” is not present in the predictive dataset. The “children” were deleted from the training dataset before the decision tree could be run on the RapidMiner. Manual sorting on Microsoft excel helps distill the final result from the processed result.

Even though these data were processed, and the final answer is gotten, the ability to pay back the loan is not one of the attributes. This particularly makes me ambivalent about which 5 records to delete out of the 25 to prune it to the required 20. It was a choice between old age and a small salary. If this had been included in the data, it could have been easy to fathom whether old people are reluctant to pay their loans or people with little income. This is because, at the end of the day, it is not about customers accepting loans as a result of a direct marketing campaign, their ability to repay the loan after they have been awarded should equally be a factor. I encourage further research into the characteristic of bank customers who easily repay their loans. This will help to accurately calculate the expected returns to prevent a situation where the expected return will not turn to expected loss after a successful target campaign.

References

Dawson, P., & Andriopoulos, C. (2017). Managing Change, Creativity and Innovation (3rd ed.). SAGE Publications Ltd.

Decision Tree Analysis: Choosing by Projecting "Expected Outcomes." (2020). Mind Tools. https://www.mindtools.com/dectree.html#:%7E:text=Decision%20trees%20provide%20an%20effective,the%20probabilities%20of%20achieving%20them

Kaplan, R.S. and Norton, D.P. (1992) The balanced scorecard – measures that drive performance. Harvard Business Review 70 (1): 71–9.

Landis, T. (2020). Customer Retention Marketing vs. Customer Acquisition Marketing. OutboundEngine. https://www.outboundengine.com/blog/customer-retention-marketing-vs-customer-acquisition-marketing/

Market Segmentation: Everything to Know in 2020 //. (2020, October 14). Qualtrics. https://www.qualtrics.com/experience-management/brand/what-is-market-segmentation/

Predictive Analytics: What it is and why it matters. (2020). SAS. https://www.sas.com/en_us/insights/analytics/predictive-analytics.html#:%7E:text=Predictive%20analytics%20are%20used%20to,forecast%20inventory%20and%20manage%20resources.

Press, G. (2016, March 23). Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data

Science Task, Survey Says. Retrieved from https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/#722588936f63

Target Market (2020). Wikipedia. https://en.wikipedia.org/wiki/Target_market

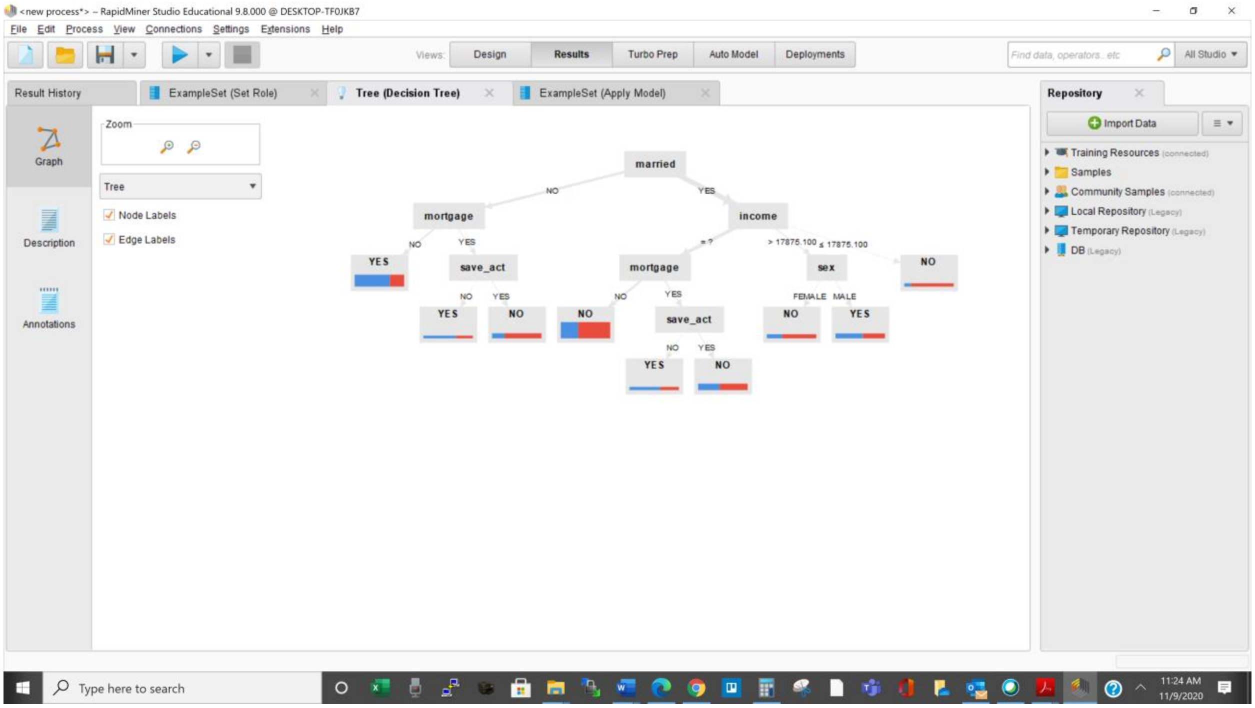
University of Eastern Finland. (n.d.). Machine learning. Retrieved from http://www.cs.joensuu.fi/~whamalai/skc/ml.html

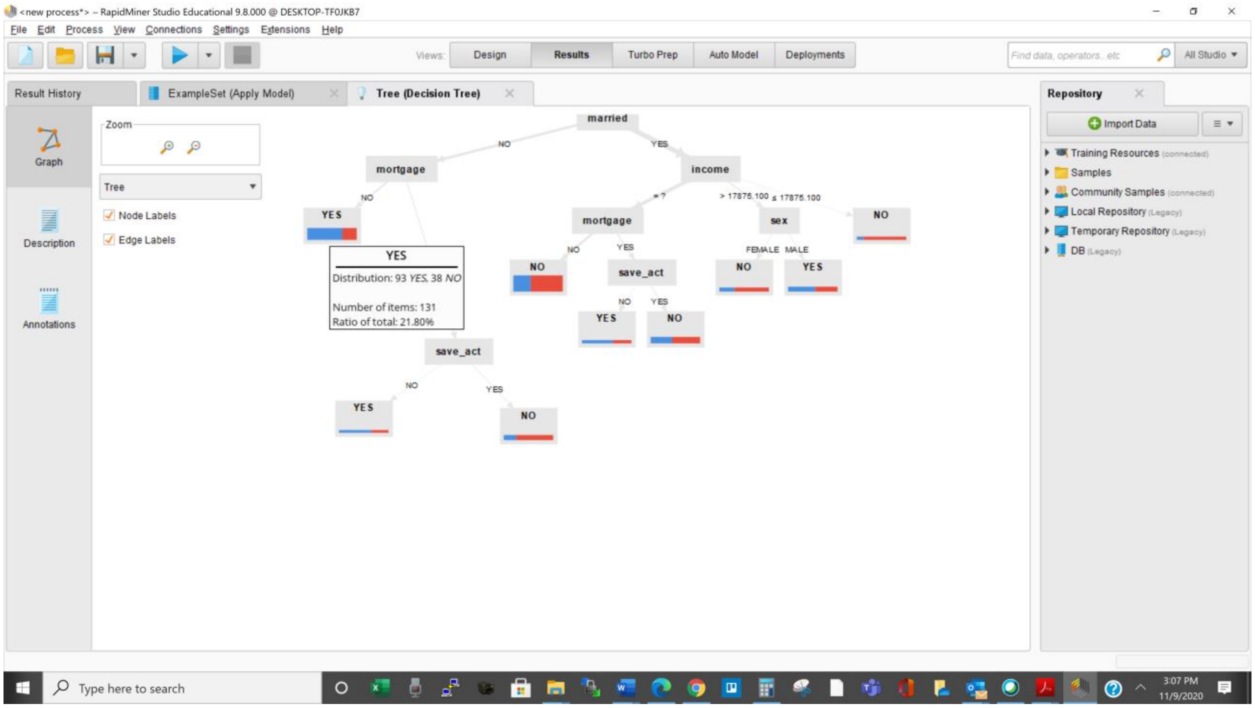
Wikipedia contributors. (2020, October 29). *Market segmentation*. Wikipedia. https://en.wikipedia.org/wiki/Market_segmentation

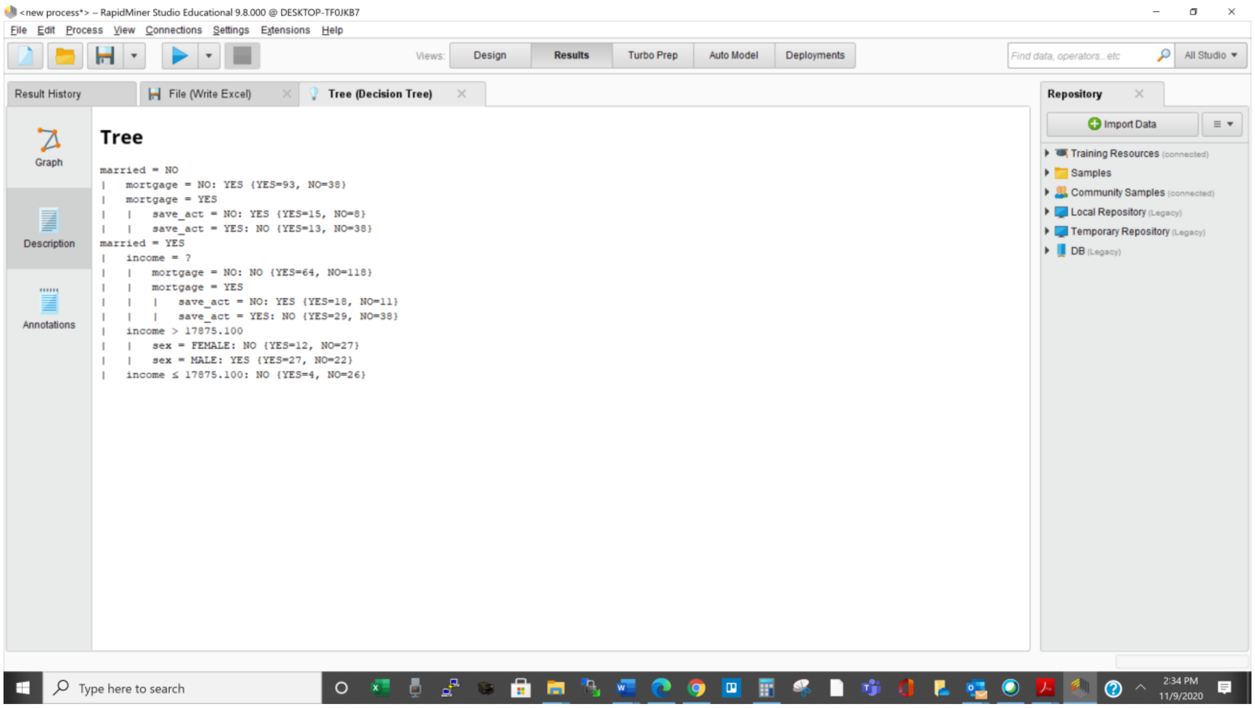
Appendices

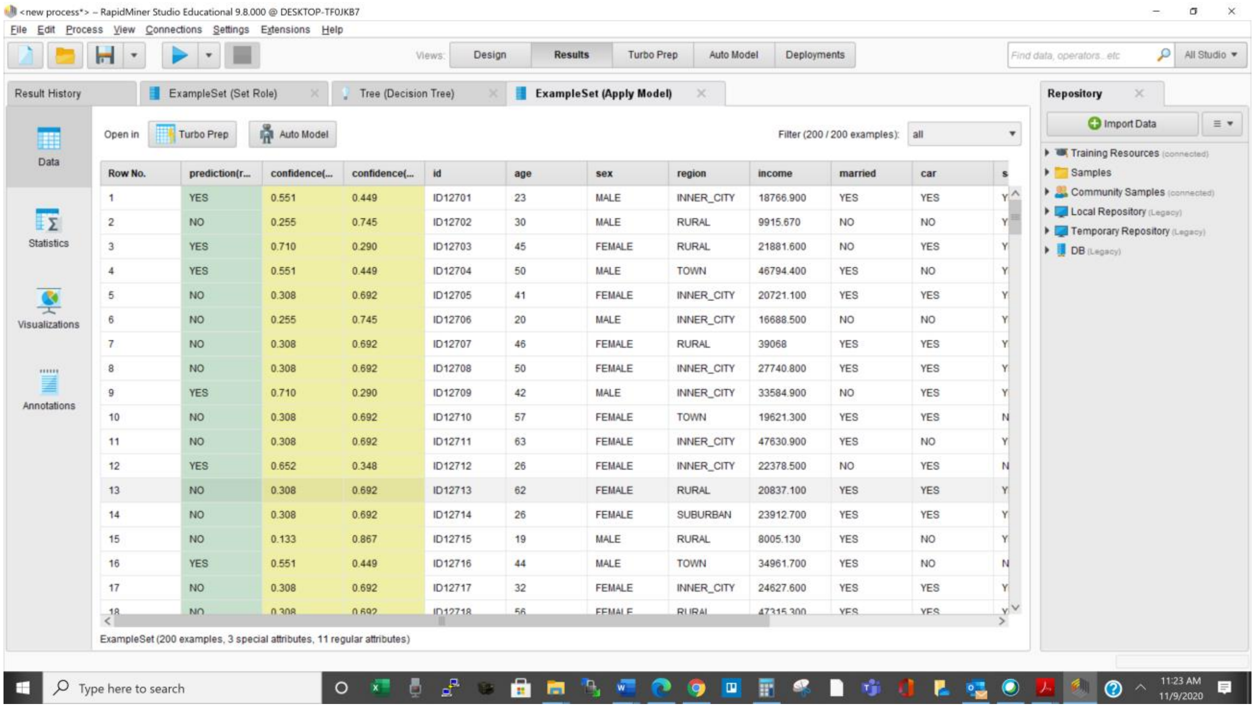
Appendix 1.

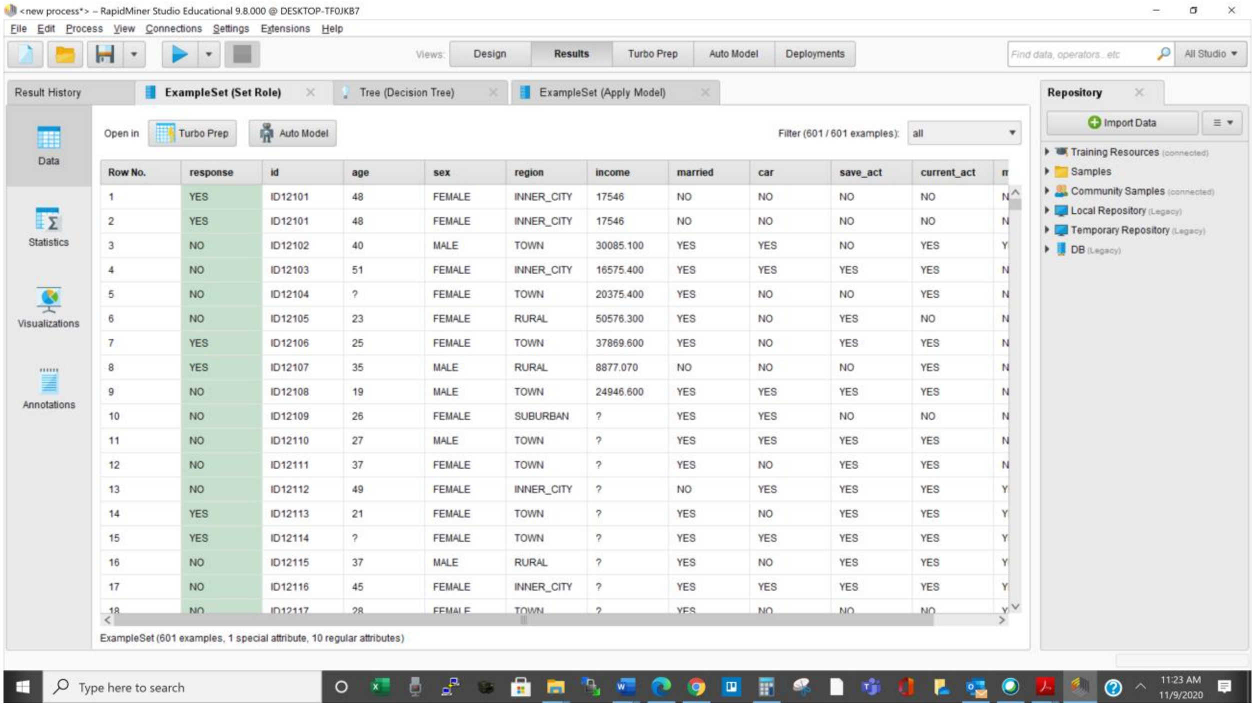
Using response as attribute name for a label





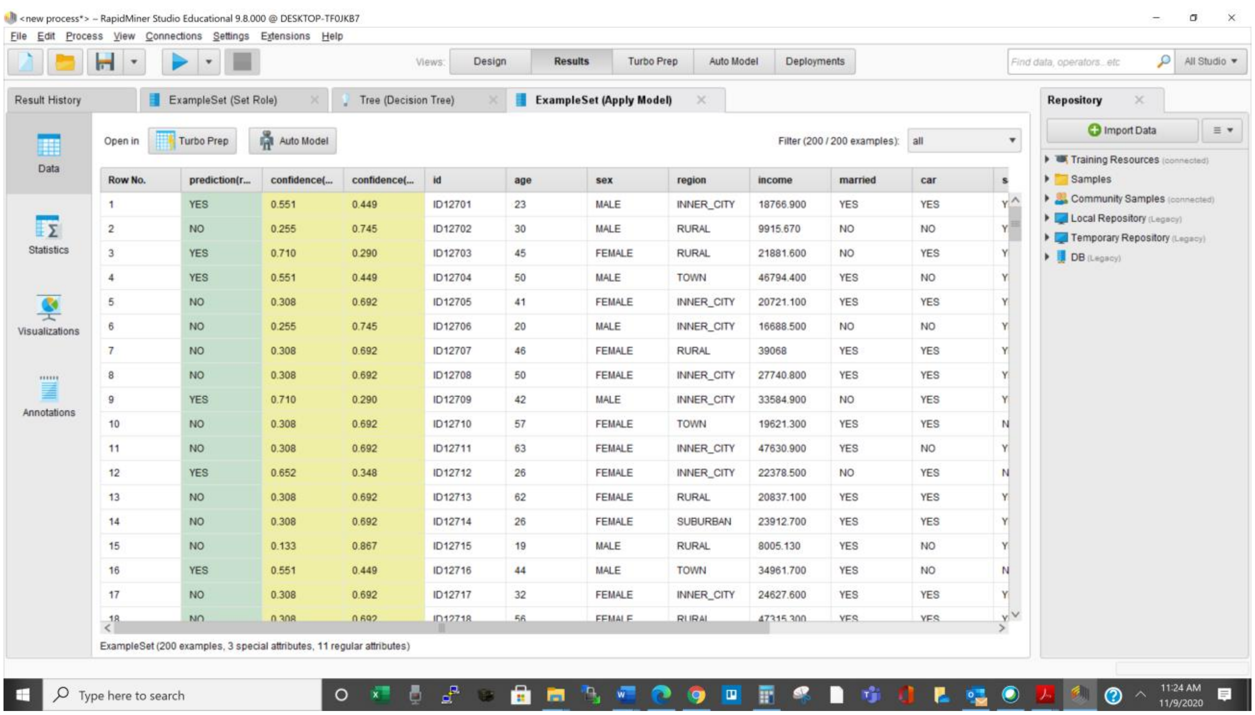

<new process*> – RapidMiner Studio Educational 9.8.000 @ DESKTOP-TF0JKB7
File Edit Process View Connections Settings Extensions Help
Views: Design Results Turbo Prep Auto Model Deployments
Result History
File (Write Excel)
Tree (Decision Tree)
Graph
Description
Annotations
Tree
married = NO
| mortgage = NO: YES {YES=93, NO=38}
| mortgage = YES
| | save_act = NO: YES {YES=15, NO=8}
| | save_act = YES: NO {YES=13, NO=38}
married = YES
| income = ?
| | mortgage = NO: NO {YES=64, NO=118}
| | mortgage = YES
| | | save_act = NO: YES {YES=18, NO=11}
| | | save_act = YES: NO {YES=29, NO=38}
| income > 17875.100
| | sex = FEMALE: NO {YES=12, NO=27}
| | sex = MALE: YES {YES=27, NO=22}
| income ≤ 17875.100: NO {YES=4, NO=26}
Repository
Import Data
Training Resources (connected)
Samples
Community Samples (connected)
Local Repository (Legacy)
Temporary Repository (Legacy)
DB (Legacy)
Type here to search
2:34 PM
11/9/2020


<new process*> – RapidMiner Studio Educational 9.8.000 @ DESKTOP-TF0JKB7
File Edit Process View Connections Settings Extensions Help
Views: Design Results Turbo Prep Auto Model Deployments
Result History
ExampleSet (Set Role)
Tree (Decision Tree)
ExampleSet (Apply Model)
Data
Statistics
Visualizations
Annotations
Open in Turbo Prep Auto Model
Filter (200 / 200 examples): all
Row No. prediction(r... confidence(... confidence(... id age sex region income married car
1 YES 0.551 0.449 ID12701 23 MALE INNER_CITY 18766.900 YES YES
2 NO 0.255 0.745 ID12702 30 MALE RURAL 9915.670 NO NO
3 YES 0.710 0.290 ID12703 45 FEMALE RURAL 21881.600 NO YES
4 YES 0.551 0.449 ID12704 50 MALE TOWN 46794.400 YES NO
5 NO 0.308 0.692 ID12705 41 FEMALE INNER_CITY 20721.100 YES YES
6 NO 0.255 0.745 ID12706 20 MALE INNER_CITY 16688.500 NO NO
7 NO 0.308 0.692 ID12707 46 FEMALE RURAL 39068 YES YES
8 NO 0.308 0.692 ID12708 50 FEMALE INNER_CITY 27740.800 YES YES
9 YES 0.710 0.290 ID12709 42 MALE INNER_CITY 33584.900 NO YES
10 NO 0.308 0.692 ID12710 57 FEMALE TOWN 19621.300 YES YES
11 NO 0.308 0.692 ID12711 63 FEMALE INNER_CITY 47630.900 YES NO
12 YES 0.652 0.348 ID12712 26 FEMALE INNER_CITY 22378.500 NO YES
13 NO 0.308 0.692 ID12713 62 FEMALE RURAL 20837.100 YES YES
14 NO 0.308 0.692 ID12714 26 FEMALE SUBURBAN 23912.700 YES YES
15 NO 0.133 0.867 ID12715 19 MALE RURAL 8005.130 YES NO
16 YES 0.551 0.449 ID12716 44 MALE TOWN 34961.700 YES NO
17 NO 0.308 0.692 ID12717 32 FEMALE INNER_CITY 24627.600 YES YES
ExampleSet (200 examples, 3 special attributes, 11 regular attributes)
Repository
Import Data
Training Resources (connected)
Samples
Community Samples (connected)
Local Repository (Legacy)
Temporary Repository (Legacy)
DB (Legacy)
Type here to search
11:23 AM
11/9/2020



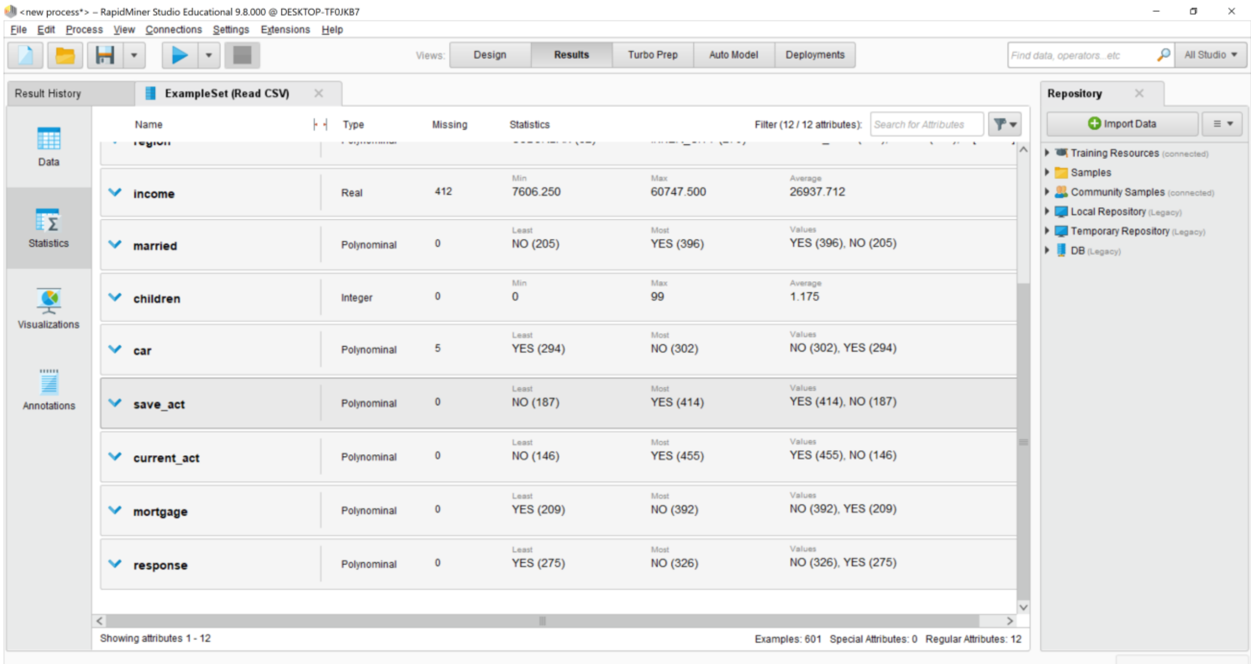
Row No.	response	id	age	sex	region	income	married	car	save_act	current_act
1	YES	ID12101	48	FEMALE	INNER_CITY	17546	NO	NO	NO	NO
2	YES	ID12101	48	FEMALE	INNER_CITY	17546	NO	NO	NO	NO
3	NO	ID12102	40	MALE	TOWN	30085.100	YES	YES	NO	YES
4	NO	ID12103	51	FEMALE	INNER_CITY	16575.400	YES	YES	YES	YES
5	NO	ID12104	?	FEMALE	TOWN	20375.400	YES	NO	NO	YES
6	NO	ID12105	23	FEMALE	RURAL	50576.300	YES	NO	YES	NO
7	YES	ID12106	25	FEMALE	TOWN	37869.600	YES	NO	YES	YES
8	YES	ID12107	35	MALE	RURAL	8877.070	NO	NO	NO	YES
9	NO	ID12108	19	MALE	TOWN	24946.600	YES	YES	YES	YES
10	NO	ID12109	26	FEMALE	SUBURBAN	?	YES	YES	NO	NO
11	NO	ID12110	27	MALE	TOWN	?	YES	YES	YES	YES
12	NO	ID12111	37	FEMALE	TOWN	?	YES	NO	YES	YES
13	NO	ID12112	49	FEMALE	INNER_CITY	?	NO	YES	YES	YES
14	YES	ID12113	21	FEMALE	TOWN	?	YES	NO	YES	YES
15	YES	ID12114	?	FEMALE	TOWN	?	YES	YES	YES	YES
16	NO	ID12115	37	MALE	RURAL	?	YES	NO	YES	YES
17	NO	ID12116	45	FEMALE	INNER_CITY	?	YES	YES	YES	YES
18	NO	ID12117	28	FEMALE	TOWN	?	YES	NO	NO	NO

ExampleSet (601 examples, 1 special attribute, 10 regular attributes)

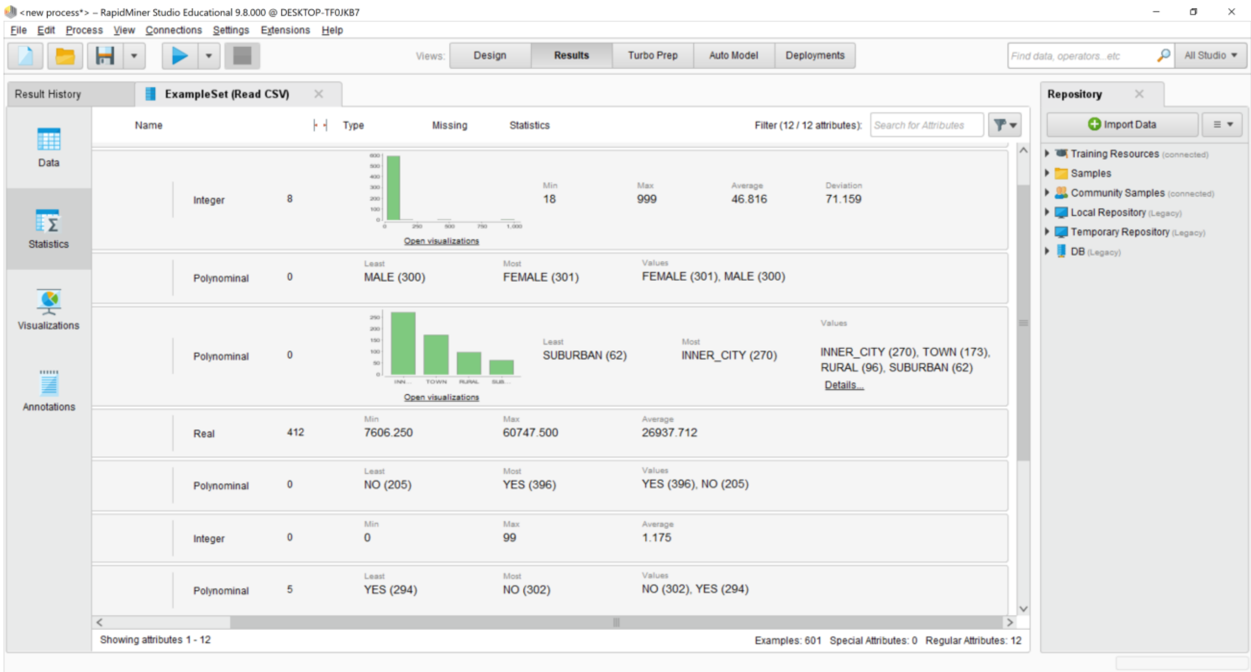


Row No.	prediction(r...	confidence(...	confidence(...	id	age	sex	region	income	married	car
1	YES	0.551	0.449	ID12701	23	MALE	INNER_CITY	18766.900	YES	YES
2	NO	0.255	0.745	ID12702	30	MALE	RURAL	9915.670	NO	NO
3	YES	0.710	0.290	ID12703	45	FEMALE	RURAL	21881.600	NO	YES
4	YES	0.551	0.449	ID12704	50	MALE	TOWN	46794.400	YES	NO
5	NO	0.308	0.692	ID12705	41	FEMALE	INNER_CITY	20721.100	YES	YES
6	NO	0.255	0.745	ID12706	20	MALE	INNER_CITY	16688.500	NO	NO
7	NO	0.308	0.692	ID12707	46	FEMALE	RURAL	39068	YES	YES
8	NO	0.308	0.692	ID12708	50	FEMALE	INNER_CITY	27740.800	YES	YES
9	YES	0.710	0.290	ID12709	42	MALE	INNER_CITY	33584.900	NO	YES
10	NO	0.308	0.692	ID12710	57	FEMALE	TOWN	19621.300	YES	YES
11	NO	0.308	0.692	ID12711	63	FEMALE	INNER_CITY	47630.900	YES	NO
12	YES	0.652	0.348	ID12712	26	FEMALE	INNER_CITY	22378.500	NO	YES
13	NO	0.308	0.692	ID12713	62	FEMALE	RURAL	20837.100	YES	YES
14	NO	0.308	0.692	ID12714	26	FEMALE	SUBURBAN	23912.700	YES	YES
15	NO	0.133	0.867	ID12715	19	MALE	RURAL	8005.130	YES	NO
16	YES	0.551	0.449	ID12716	44	MALE	TOWN	34961.700	YES	NO
17	NO	0.308	0.692	ID12717	32	FEMALE	INNER_CITY	24627.600	YES	YES
18	NO	0.308	0.692	ID12718	56	FEMALE	RURAL	47315.300	YES	YES

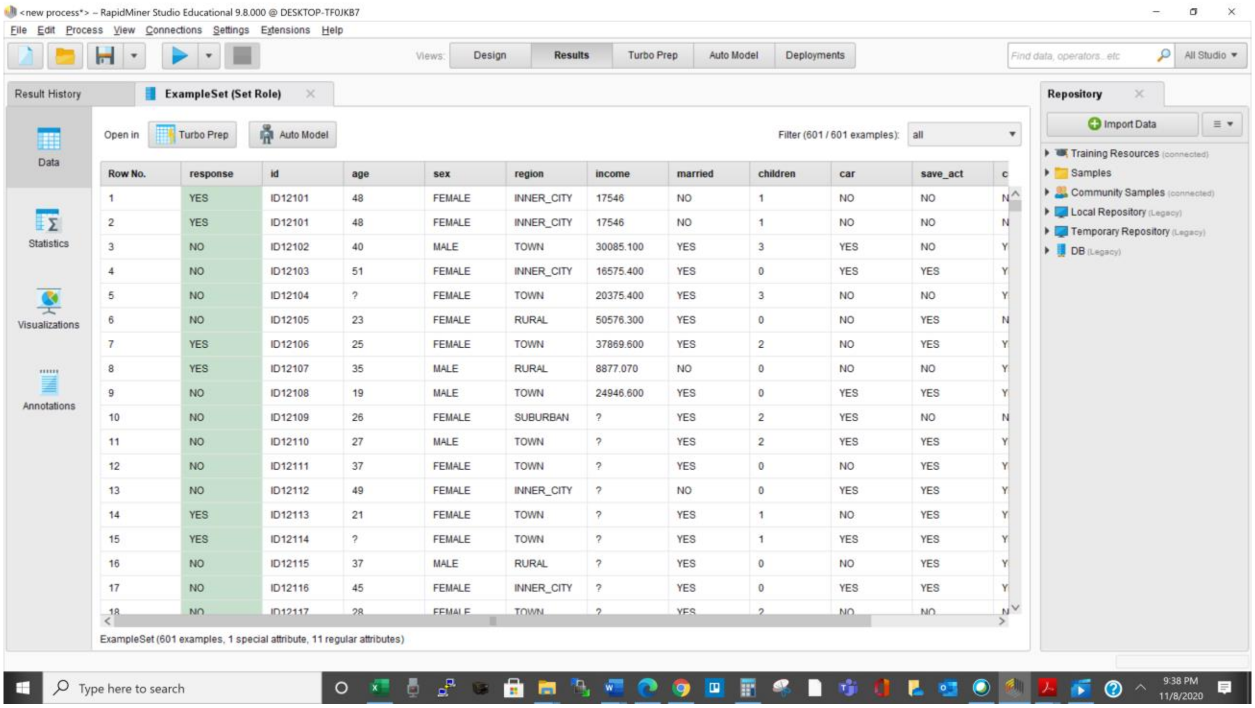
ExampleSet (200 examples, 3 special attributes, 11 regular attributes)



<new process*> – RapidMiner Studio Educational 9.8.000 @ DESKTOP-TF0JKB7
File Edit Process View Connections Settings Extensions Help
Views: Design Results Turbo Prep Auto Model Deployments
Result History
ExampleSet (Read CSV)
Repository
Import Data
Data
Statistics
Visualizations
Annotations
Name Type Missing Statistics Filter (12 / 12 attributes):
income Real 412 Min 7606.250 Max 60747.500 Average 26937.712
married Polynominal 0 Least NO (205) Most YES (396) Values YES (396), NO (205)
children Integer 0 Min 0 Max 99 Average 1.175
car Polynominal 5 Least YES (294) Most NO (302) Values NO (302), YES (294)
save_act Polynominal 0 Least NO (187) Most YES (414) Values YES (414), NO (187)
current_act Polynominal 0 Least NO (146) Most YES (455) Values YES (455), NO (146)
mortgage Polynominal 0 Least YES (209) Most NO (392) Values NO (392), YES (209)
response Polynominal 0 Least YES (275) Most NO (326) Values NO (326), YES (275)
Training Resources
Samples
Community Samples
Local Repository
Temporary Repository
DB
Showing attributes 1 - 12
Examples: 601 Special Attributes: 0 Regular Attributes: 12

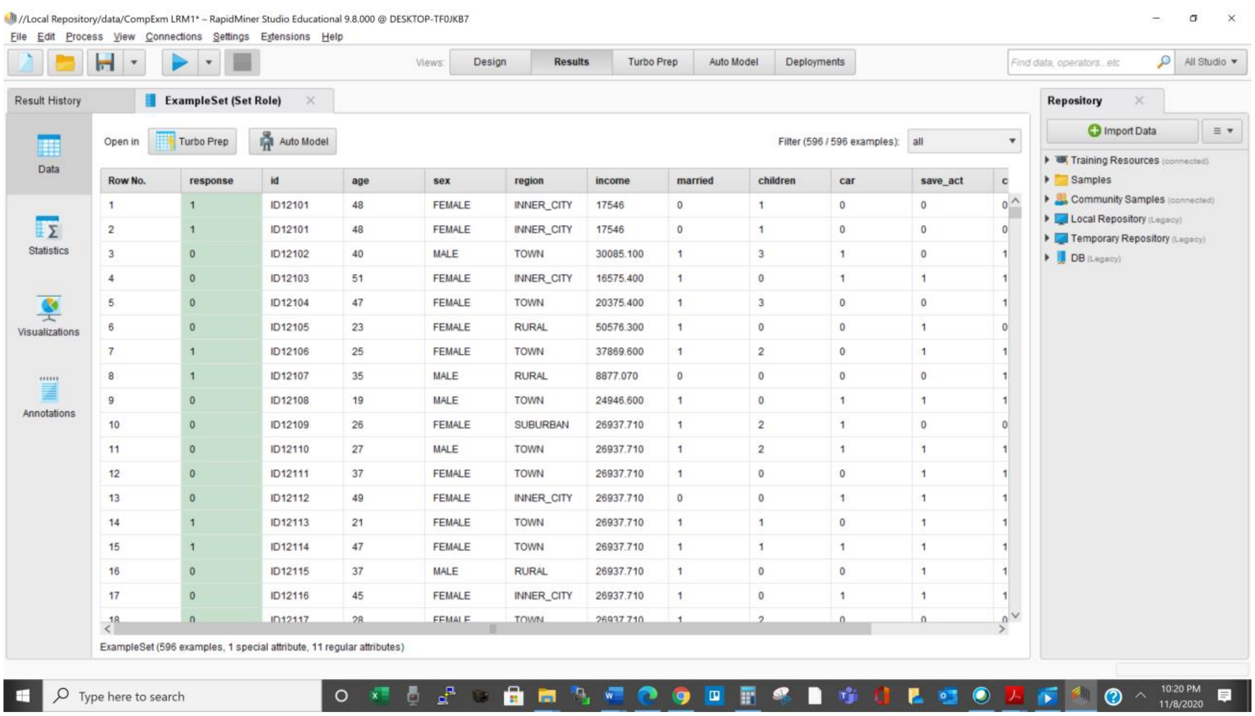


<new process*> – RapidMiner Studio Educational 9.8.000 @ DESKTOP-TF0JKB7
File Edit Process View Connections Settings Extensions Help
Views: Design Results Turbo Prep Auto Model Deployments
Result History
ExampleSet (Read CSV)
Repository
Import Data
Name Type Missing Statistics Filter (12 / 12 attributes):
Integer 8 Min 18 Max 999 Average 46.816 Deviation 71.159
Open visualizations
Polynominal 0 Least MALE (300) Most FEMALE (301) Values FEMALE (301), MALE (300)
Polynominal 0 Least SUBURBAN (62) Most INNER_CITY (270) Values INNER_CITY (270), TOWN (173), RURAL (96), SUBURBAN (62)
Details...
Open visualizations
Real 412 Min 7606.250 Max 60747.500 Average 26937.712
Polynominal 0 Least NO (205) Most YES (396) Values YES (396), NO (205)
Integer 0 Min 0 Max 99 Average 1.175
Polynominal 5 Least YES (294) Most NO (302) Values NO (302), YES (294)
Training Resources
Samples
Community Samples
Local Repository
Temporary Repository
DB
Showing attributes 1 - 12
Examples: 601 Special Attributes: 0 Regular Attributes: 12



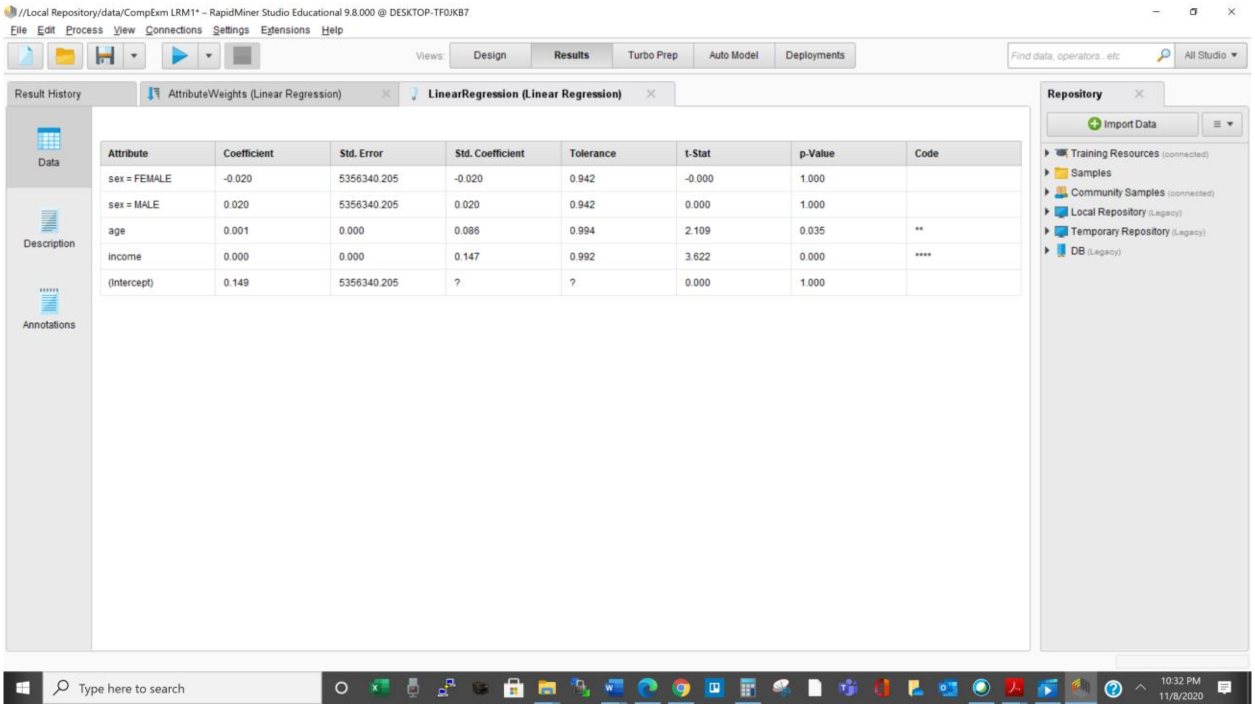
Row No.	response	id	age	sex	region	income	married	children	car	save_act	c
1	YES	ID12101	48	FEMALE	INNER_CITY	17546	NO	1	NO	NO	N
2	YES	ID12101	48	FEMALE	INNER_CITY	17546	NO	1	NO	NO	N
3	NO	ID12102	40	MALE	TOWN	30085.100	YES	3	YES	NO	Y
4	NO	ID12103	51	FEMALE	INNER_CITY	16575.400	YES	0	YES	YES	Y
5	NO	ID12104	?	FEMALE	TOWN	20375.400	YES	3	NO	NO	Y
6	NO	ID12105	23	FEMALE	RURAL	50576.300	YES	0	NO	YES	N
7	YES	ID12106	25	FEMALE	TOWN	37869.600	YES	2	NO	YES	Y
8	YES	ID12107	35	MALE	RURAL	8877.070	NO	0	NO	NO	Y
9	NO	ID12108	19	MALE	TOWN	24946.600	YES	0	YES	YES	Y
10	NO	ID12109	26	FEMALE	SUBURBAN	?	YES	2	YES	NO	N
11	NO	ID12110	27	MALE	TOWN	?	YES	2	YES	YES	Y
12	NO	ID12111	37	FEMALE	TOWN	?	YES	0	NO	YES	Y
13	NO	ID12112	49	FEMALE	INNER_CITY	?	NO	0	YES	YES	Y
14	YES	ID12113	21	FEMALE	TOWN	?	YES	1	NO	YES	Y
15	YES	ID12114	?	FEMALE	TOWN	?	YES	1	YES	YES	Y
16	NO	ID12115	37	MALE	RURAL	?	YES	0	NO	YES	Y
17	NO	ID12116	45	FEMALE	INNER_CITY	?	YES	0	YES	YES	Y
18	NO	ID12117	28	FEMALE	TOWN	?	YES	2	NO	NO	N

ExampleSet (601 examples, 1 special attribute, 11 regular attributes)



Row No.	response	id	age	sex	region	income	married	children	car	save_act	c
1	1	ID12101	48	FEMALE	INNER_CITY	17546	0	1	0	0	0
2	1	ID12101	48	FEMALE	INNER_CITY	17546	0	1	0	0	0
3	0	ID12102	40	MALE	TOWN	30085.100	1	3	1	0	1
4	0	ID12103	51	FEMALE	INNER_CITY	16575.400	1	0	1	1	1
5	0	ID12104	47	FEMALE	TOWN	20375.400	1	3	0	0	1
6	0	ID12105	23	FEMALE	RURAL	50576.300	1	0	0	1	0
7	1	ID12106	25	FEMALE	TOWN	37869.600	1	2	0	1	1
8	1	ID12107	35	MALE	RURAL	8877.070	0	0	0	0	1
9	0	ID12108	19	MALE	TOWN	24946.600	1	0	1	1	1
10	0	ID12109	26	FEMALE	SUBURBAN	26937.710	1	2	1	0	0
11	0	ID12110	27	MALE	TOWN	26937.710	1	2	1	1	1
12	0	ID12111	37	FEMALE	TOWN	26937.710	1	0	0	1	1
13	0	ID12112	49	FEMALE	INNER_CITY	26937.710	0	0	1	1	1
14	1	ID12113	21	FEMALE	TOWN	26937.710	1	1	0	1	1
15	1	ID12114	47	FEMALE	TOWN	26937.710	1	1	1	1	1
16	0	ID12115	37	MALE	RURAL	26937.710	1	0	0	1	1
17	0	ID12116	45	FEMALE	INNER_CITY	26937.710	1	0	1	1	1
18	0	ID12117	28	FEMALE	TOWN	26937.710	1	2	0	0	0

ExampleSet (596 examples, 1 special attribute, 11 regular attributes)



Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value	Code
sex = FEMALE	-0.020	5356340.205	-0.020	0.942	-0.000	1.000	
sex = MALE	0.020	5356340.205	0.020	0.942	0.000	1.000	
age	0.001	0.000	0.086	0.994	2.109	0.035	**
income	0.000	0.000	0.147	0.992	3.622	0.000	****
(Intercept)	0.149	5356340.205	?	?	0.000	1.000