DATA MINING

Department of Computer Science and Engineering
National Institute of Technology Calicut, Kozhikode.

Date: 24th April 2015

Abstract

This report outlines the contents for documenting the work done as - Machine Learning project. The objective of the following project was to predict if a car purchased at an auction is a kick (bad buy). I was provided with a messy data set containing 72983 tuples and 37 attributes. Initially, I applied data cleaning techniques to improve the quality of the data set. Thereafter, I used appropriate mining techniques to arrive at the desired conclusions. Using a decision tree model, with the given ExampleSet split into 70% training set and 30% test set, I was able to get an accuracy of 88.04%.

Contents of the Documentation set

1. Introduction
    a. Project Overview
    b. Project Deliverables
2. Project Organization
    a. Process Model
    b. Tools and Techniques Used

I. Project Management Plan
    a) Tasks
    b) Description of the plan
    c) Data set Description
    d) Deliverables and Milestones- give details
    e) Dependencies and other constraints

II. Requirement Specification
    a. Hardware Requirement
    b. Specific Requirements
        i. User Interface
        ii. Software Interface
        iii. Other details regarding the tool if any.
    c. About the Software: For Cleaning and Mining – write separately.
        i. Introduction
        ii. Reliability
        iii. Availability
        iv. Security
        v. Maintainability
        vi. Portability
        vii. Performance

III. Data base requirements

IV. Design Specifications
    a. Work Done
        i. Chosen tool set: Pros and Cons
        ii. Correlation
        iii. Classification Models
        iv. Clustering Model
        v. Association and Correlation Model
        vi. Data Visualization

V. Test Documentation
    a. Features to be tested
    b. Test cases
        i. Case n
            1. Purpose
            2. Input
            3. Expected output
            4. Test procedure
    c. Test Logs

1. Introduction

a. Project Overview

One of the biggest challenges of an auto dealership purchasing a used car at an auto auction is the risk of that the vehicle might have serious issues that prevent it from being sold to customers. The auto community calls these unfortunate purchases "kicks".

Kicked cars often result when there are tampered odometers, mechanical issues the dealer is not able to address, issues with getting the vehicle title from the seller, or some other unforeseen problem. Kick cars can be very costly to dealers after transportation cost, throw-away repair work, and market losses in reselling the vehicle.

Modellers who can figure out which cars have a higher risk of being kick can provide real value to dealerships trying to provide the best inventory selection possible to their customers.

The challenge of this competition is to predict if the car purchased at the Auction is a Kick (bad buy).

To find out the information , I have analyzed and cleaned the data set containing various car auction details during the period January, 2009 to December, 2010.

b. Project Deliverables

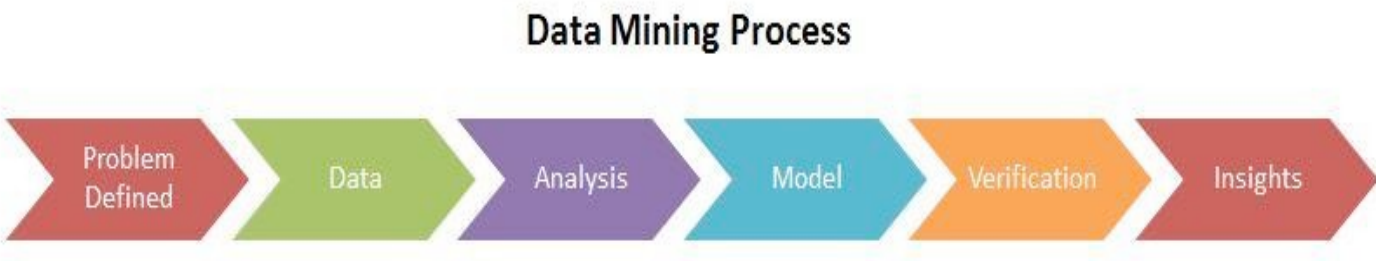
Data Cleaning

The concepts discussed in theory sessions which were applied while cleaning were:

1. Use of measure of central tendency such as the mean or median method for filling in the missing values. Measures of central tendency indicate the “middle” value of a data distribution. For normal(symmetric) data distribution, the mean can be used, while skewed data distribution should employ the median or mode.
2. Use of the most probable value to fill in the missing value. This may be determined with regression, inference based tools such as decision tree indication.
3. Fill in the missing data manually for tuples with minor errors like spelling mistakes occurring in low frequency.

2. 2. Project Organization

a. Process Model



- Problem Definition : Understand the project objectives and requirements from a business perspective, and then convert this knowledge into a data mining problem definition and a preliminary plan designed to achieve the objectives.
- Data Understanding : Start by collecting data, then get familiar with the data, to identify data quality problems, to discover first insights into the data, or to detect interesting subsets to form hypotheses about hidden information.
- Data Analysis : Includes all activities required to construct the final data set (data that will be fed into the modeling tool) from the initial raw data. Tasks include table, case, and attribute selection as well as transformation and cleaning of data for modeling tools.
- Modeling : Select and apply a variety of modelling techniques, and calibrate tool parameters to optimal values. Typically, there are several techniques for the same data mining problem type. Some techniques have specific requirements on the form of data. Therefore, stepping back to the data preparation phase is often needed.
- Data Verification : It is a process where in different types of data are checked for accuracy and inconsistencies after data migration is done. It helps to determine whether data was accurately translated when data is transferred from one source to another, is complete, and supports processes in the new system.
- Evaluation and Insights : Thoroughly evaluate the model, and review the steps executed to construct the model, to be certain it properly achieves the business objectives. Determine if there is some important business issue that has not been sufficiently considered. At the end of this phase, a decision on the use of the data mining results is reached.

b. Tools and Techniques Used

The tool used for data cleaning was OpenRefine Version 2.5 .

OpenRefine is a standalone open source desktop application for data cleanup and transformation to other formats, the activity known as data wrangling. It is similar to spreadsheet applications (and can work with spreadsheet file formats), however, it behaves more like a database.

It operates on rows of data which have cells under columns, which is very similar to relational database tables. One OpenRefine project is one table. The program has a web user interface. When starting OpenRefine, it starts a web server and starts a browser to open the web UI powered by this web server.

The tool used for data mining was Rapid Miner.

RapidMiner is a software platform developed by the company of the same name that provides an integrated environment for machine learning, data mining, text mining, predictive analytics and business analytics.It is used for business and industrial applications as well as for research, education, training, rapid prototyping, and application development and supports all steps of the data mining process including results visualization, validation and optimization.[1] RapidMiner is developed on a business source model which means the core and earlier versions of the software are available under an OSI-certified open source license on Sourceforge.

I. Project Management Plan

a. Tasks

Section 1 : Data Cleaning

There are many features in OpenRefine. I have focussed on the most used and important features of OpenRefine. They are listed as follows:

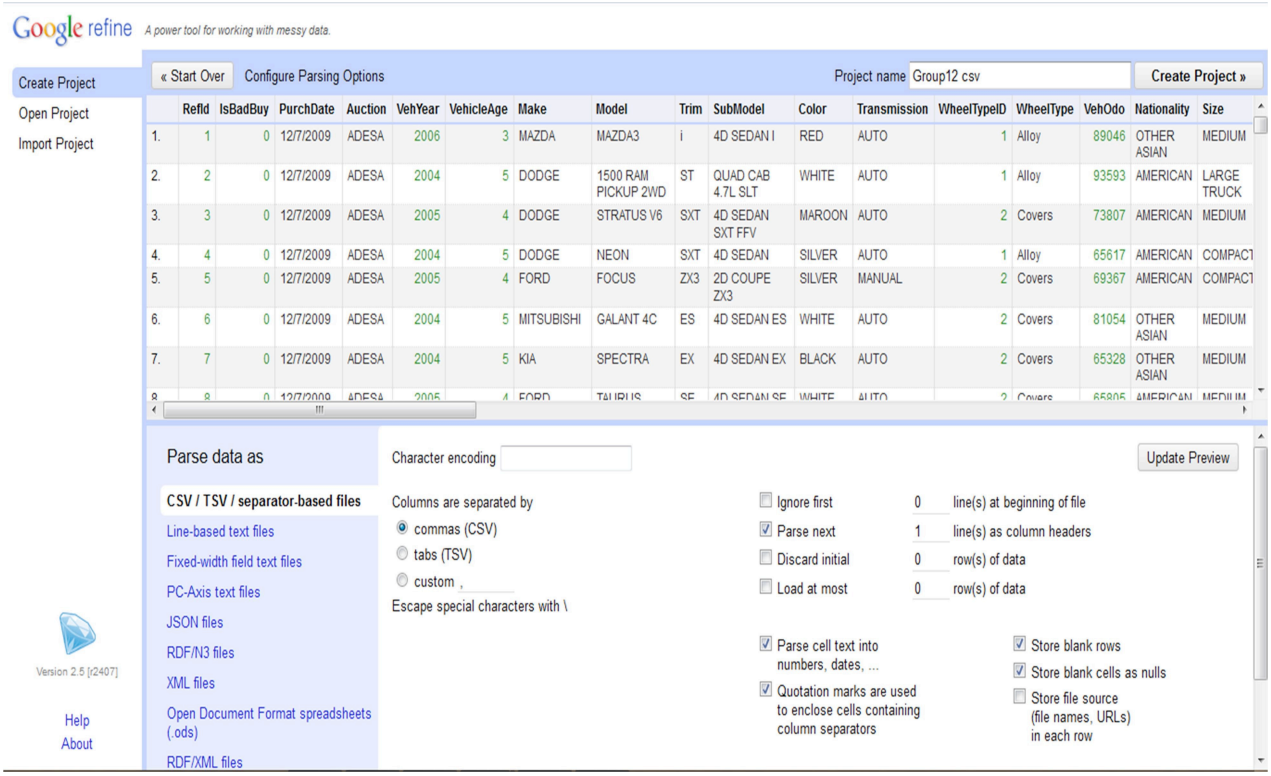
a. Importing Data
b. Creation of Data
c. Filtering/ Faceting
d. Editing cells/columns
e. Exporting Data
f. Undo/ Redo

A brief explanation of the features

a) Importing Data :

The importing data is used to get the data from various external sources.

b) Creation of Project :



c) Filtering/ Faceting

RefId is a unique number assigned to each auctioned item. It was observed that RefId wasn't strictly sequential (incremental gaps between two ID's at random places). A numeric facet was applied and it was found that all the values in the column appeared to be numeric and whole numbers. The ID's were in a sorted order.

PurchDate was checked for inconsistencies with a timeline facet and was found to fall in the range of Jan 2009 to December 2010 using a text transformation, text facet and a sort. This attribute did not contain any blanks or NULL's.

Auction was checked with a text facet as well. Consisting of three main divisions namely - ADESA, MANHEIM and OTHERS. This column did not require any editing as no inconsistencies turned up on a facet. A common transformation to uppercase was done.

VehYear and VehicleAge did not require any cleaning. But the VehYear was in the text format and required to be transformed to numbers. This was accomplished with the help of the common transformations feature using the 'To number' transformation. VehicleAge was found to be clean with the numeric facet and since all the cells had numeric values, it did not need a transformation.

Make, Model and SubModel were the attributes that needed most cleaning. A common transformation of 'To text', 'To uppercase' and 'Trim leading and trailing whitespace' was performed in accordance with the metadata.

Trim was capitalized with the common transformations – 'To text' and 'To uppercase'. NULL's and blanks were found on text faceting. Necessary filtering using attributes – Make, Model, SubModel, Color and Size of the car was used to fill in the missing values and remove the NULL (replacement with most probable value).
Color was converted to text and capitalised using necessary transformations. Text faceting showed the presence of NULL values and blanks as well. Filtering based on the auction date, make and model were done to fill in the NULL and blanks.

Transmission consisted of NULL, blank and a mix of lower case & uppercase text. Accordingly, an uppercase transformation was applied and the NULL and blank values were filtered on the basis of attributes – Make, Size, and Model.

WheelTypeID and WheelType were found to be related. The following information was obtained from doing a facet and filtering on both attributes

a. WheelTypeID of 1 implies that the WheelType has a value of 'Alloy'.
b. WheelTypeID of 2 implies that the WheelType has a value of 'Covers'.
c. WheelTypeID of 3 implies that the WheelType has a value of 'Special'.

A numeric transformation was used and a round transformation was used as well present in the transform tab. WheelTypeID had values of 0,1,2,3 and NULL values. The tuples having a WheelTypeID of 0 were merged with tuples having the same attribute as NULL. I then obtained values for the NULL and blanks in WheelType based on attributes – Make, Model, Trim, Color and Size. After which, I filled the WheelTypeID with the mapping above.

VehOdo had the following transformations performed on it – 'round()' and 'To number'. The attribute did not provide us with any inconsistencies to clean up.

For Nationality, I applied Text facet, obtained 4 different classes and NULL values. The NULL values were replaced with appropriate filtering based on the Make attribute.

For TopThreeAmericanName, I applied a 'To uppercase' transform and Text facet, obtained 4 different classes and NULL values. The NULL values were replaced with appropriate filtering based on the Make attribute as well.

The attributes named with MMR as the initial were cleaned as follows.

- MMRAcquisitionAuctionAveragePrice
- MMRAcquisitionAuctionCleanPrice
- MMRAcquisitionRetailAveragePrice
- MMRAcquisitonRetailCleanPrice,
- MMRCurrentAuctionAveragePrice
- MMRCurrentAuctionCleanPrice,
- MMRCurrentRetailAveragePrice
- MMRCurrentRetailCleanPrice

A round transformation was used on all the attributes since the prices had to be reported in whole numbers according to the Manheim Market Report(MMR). NULL values were obtained. Inorder to fill these up, I used mean aggregrate functions and replaced the NULL with the mean of their respective columns.

BYRNO was initially transformed using a 'To number' and then a numeric facet was applied. No NULL's, blanks or non-numeric values were present.

VNZIP1 and VNST are related attributes. They represent the US state codes and the abbreviations respectively. 37 states of the US have been represented as obtained using a text facet and each of them were checked using the following reference site - http://en.wikipedia.org/wiki/List_of_U.S._state_abbreviations and found to be accurate. The issue encountered in VNZIP1 was that certain tuples had zip numbers of only 4 digits when 5 digits was mandatory. This was observed when a Text Length Facet was performed. This was corrected by applying a text facet on VNST and VNZIP1 and by selecting the 4 digit tuples in VNZIP1 and prefixing the values with the appropriate state code(first digit).

VehBCost was worked on with a round transform function followed by a 'To number' transform. No NULL's, blanks and non-numeric values were obtained.

IsOnlineSale is a binary attribute consisting of just 1's and 0's. A text facet showed no NULL's, blanks or any other values other than ones specified above.

WarrantyCost attribute was rounded using the transformation given. A numeric facet showed no non-numeric values, NULL's or other inconsistencies.

Section 2 : Data Mining

d) The partially cleaned data was converted to CSV. This was loaded into RapidMiner. The Attribute Value Type was set for all attributes. After that NULL was declared to be missing value for nominal data type. The attributes WheelType, PRIMEUNIT were split into sets based on whether the value was missing or not. Then for the missing data the value was predicted using a Decision Tree for WheelType and Naive Bayesian Classifier for PRIMEUNIT. AUCGUART was cleaned by splitting into two sets, one with missing values and a decision tree was trained without pruning or prepruning. The values were predicted. For all the locations where the predicted values were NULL, I replaced it with YELLOW.

b. Description of the plan

The project aims to check the accuracy of the isBadBuy attribute. Prior to the hands on cleaning, concepts in data preprocessing such as redundancy checks, filling missing values and checking for inconsistencies are to be brought forward. After analyzing the metadata(Caravan Data Dictionary) I read and understand the different concepts associated with the domain of the data set. After gaining much insight, I narrowed down the cleaning techniques to be used for each attribute and relationship between attributes if any. Our data set consisted of 72983 rows and 37 attributes. It was decided I will utilize OpenRefine, an open source power tool for data cleaning. The different features of OpenRefine will be studied and tested on sample data sets. The data cleaning part will be handled by two members of the team. RapidMiner cleaned the PRIMEUNIT and AUCGUART attributes. Different Classification Models were tested within the limits of the hardware and the best one was chosen based on accuracy.

b. Data Set Description

Field Name	Definition
RefID	Unique (sequential) number assigned to vehicles
IsBadBuy	Identifies if the kicked vehicle was an avoidable purchase
PurchDate	The Date the vehicle was Purchased at Auction
Auction	Auction provider at which the vehicle was purchased
VehYear	The manufacturer's year of the vehicle
VehicleAge	The Years elapsed since the manufacturer's year
Make	Vehicle Manufacturer
Model	Vehicle Model
Trim	Vehicle Trim Level
SubModel	Vehicle Submodel
Color	Vehicle Color
Transmission	Vehicles transmission type (Automatic, Manual)
WheelTypeID	The type id of the vehicle wheel
WheelType	The vehicle wheel type description (Alloy, Covers)
VehOdo	The vehicles odometer reading
Nationality	The Manufacturer's country
Size	The size category of the vehicle (Compact, SUV, etc.)
TopThreeAmericanName	Identifies if the manufacturer is one of the top three American manufacturers
MMRAcquisitionAuctionAveragePrice	Acquisition price for this vehicle in average condition at time of purchase
MMRAcquisitionAuctionCleanPrice	Acquisition price for this vehicle in the above Average condition at time of purchase
MMRAcquisitionRetailAveragePrice	Acquisition price for this vehicle in the retail market in average condition at time of purchase
MMRAcquisitonRetailCleanPrice	Acquisition price for this vehicle in the retail market in above average condition at time of purchase
MMRCurrentAuctionAveragePrice	Acquisition price for this vehicle in average condition as of current day
MMRCurrentAuctionCleanPrice	Acquisition price for this vehicle in the above condition as of current day
MMRCurrentRetailAveragePrice	Acquisition price for this vehicle in the retail market in average condition as of current day
MMRCurrentRetailCleanPrice	Acquisition price for this vehicle in the retail market in above average condition as of current day
PRIMEUNIT	Identifies if the vehicle would have a higher demand than a standard purchase
AcquisitionType*	Identifies how the vehicle was aquired (Auction buy, trade in, etc)
AUCGUART	The level guarntee provided by auction for the vehicle
KickDate	Date the vehicle was kicked back to the auction
BYRNO	Unique number assigned to the buyer that purchased the vehicle
VNZIP	Zipcode where the car was purchased
VNST	State where the the car was purchased
VehBCost	Acquisition cost paid for the vehicle at time of purchase
IsOnlineSale	Identifies if the vehicle was originally purchased online
WarrantyCost	Warranty price (term=36month and millage=36K)

c. Deliverables and Milestones

Selection of the data set	Day 1
Familiarised and tested OpenRefine on sample data	Day 2
Cleaned RefID,PurchDate,Auction and VehYear	Day 3
Cleaned VehicleAge, Make, WarrantyCost and IsOnline Sale	Day 4
Cleaned VehBCost,VNST and Model partially	Day 5
Cleaned Model, Trim and SubModel partially	Day 6
Cleaned Color, Size , TopThreeAmerican, Nationality	Day 7
Cleaned VehOdo, Model and SubModel Cleaned	Day 8
VNZIP1, BYRNO and all MMR attributes Cleaned	Day 9
WheelType and WheelTypeID	Day 9

d. Dependencies and other constraints

- The project is expected to be completed by 24th April, 2015.
- The Data set contains various dependencies as highlighted:
  - VNZIP1 and VNST refer to US state codes and US state abbreviations. Hence, they must match properly.
  - Make, Model and SubModel are all interrelated for each car at the auction.
  - Size and Model are interdependent as a particular model comes in certain sizes only.
  - WheelType is dependent on the Model.
  - TopThreeAmerican and Nationality depend on Make(i.e, manufacturer).
  - VehicleAge, VehYear and PurchDate are all interrelated.
- The constraints on the data set include:
  - RefID- Unique whole number
  - PurchDate – Follows date format
  - Auction – Any of the three: ADESA, MANHEIM and OTHERS.
  - VehYear – 4 digit number representing the year
  - VehicleAge – Whole number
  - Make – Valid Car Manufacturer
  - Model – Valid car model associated with the Make
  - SubModel – Valid vehicle submodel associated with Model
  - Color – Any of the 16 colors given
  - Transmission – AUTO or MANUAL
  - WheelTypeID – Either 1,2 or 3. Must correspond to WheelType.
  - WheelType – One of the following : Alloy, Covers or Special
  - Size – Size category according to the Make, Model and SubModel
  - VNZIP1- 5 digit valid US state code
  - VNST – 2 lettered US state abbreviation
  - VehBCost – Whole Number
  - IsOnlineSale – Binary value(0 or 1)

II. Requirements Specification

a. Hardware Requirement

- A minimum of 2 gigabytes of RAM is required
- Disk space of atleast 117 megabytes is required for installation purpose.
- Compatible with i5 and i7 intel processors.
- Supported platforms include Windows, BSD, Mac and Linux.

b. Specific Requirements

- User Interface : The interface used for OpenRefine is any web browser. I worked on Mozilla Firefox. OpenRefine resembles an excel-like interface
  which has many functionalities that include common transformations, clustering and faceting
- Software Interface : RapidMiner is a software platform developed by the company of the same name that provides an integrated environment for
  machine learning, data mining, text mining, predictive analytics and business analytics.It is used for business and industrial applications as well as for research, education, training, rapid prototyping, and application development and supports all steps of the data mining process including results visualization, validation and optimization.[1] RapidMiner is developed on a business source model which means the core and earlier versions of the software are available under an OSI-certified open source license on Sourceforge.
- Other Details regarding tools: Import is supported from following format TSV, CSV, Text file with custom separators or columns split by fixed width XML, RDF triples (RDF/XML and Notation3 serialization formats), JSON, Google Spreadsheets, Google Fusion Tables, Export is supported in following formats: Microsoft Excel,HTML table

Pros:

1. Powerful due to its learning operators and operator framework, which allows to form nearly arbitrary processes
2. Results are showed also in graphic mode.
3. It supports and can accept new data drivers.
4. It's easy and friendly to use.
5. It provides wizards for data management.

Cons

1. The system kept crashing on some models when the dataset was large.

c. About the Software

Data Cleaning

i. Introduction

OpenRefine is a standalone open source desktop application for data cleanup and transformation to other formats, the activity known as data wrangling.[3] It is similar to spreadsheet applications (and can work with spreadsheet file formats), however, it behaves more like a database. The program has a web user interface. However, it is not hosted on the web (SAAS), but is available for download and use on the local machine. When starting OpenRefine, it starts a web server and starts a browser to open the web UI powered by this web server.

ii. Reliability

Web content now includes programs that are executed directly within a web browser .Executable content, though, creates new reliability problems for users who rely on the browser to provide program services typical of operating systems. In particular, I find that the runtime environments of current browsers poorly isolate applications from one another. As a result, one web application executing within the browser can interfere with others, whether it be through an explicit failure or the excessive consumption of resources. Also note the reliability depends on the quality of the data set.

iii. Availability

OpenRefine is an open source software which can be freely downloaded from http://openrefine.org/.

iv. Security

OpenRefine is a desktop application that runs on a web browser locally. It is a personal and private web application keeping all private data from being uploaded. By default (and for security reasons) Refine only listens to TCP requests coming from localhost (127.0.0.1).

v. Maintainability

OpenRefine is written in Java and is good for long term maintainablility of data. The initial release was on 10th November 2010. Currently the stable version 2.5 was released in December,2011. Development is still in active phase.

vi. Portability

Supported by the following OS: Microsoft Windows, BSD, Linux and Macintosh. Compatible browsers include Safari, Firefox, IE(8 and above), Chrome and Opera.

vii. Performance

OpenRefine is a power tool for messy data. The usage of Heap memory that limit the real use cases with huge files. The "Energy" tab of the Activity Monitor also shows that "Requires High Performance GPU" is "Yes" for OpenRefine.

Data Mining

i. Introduction

RapidMiner is a data mining program that manages information and provides useful information about it according to parameters selected during the process setup. Such process is integrated by two components: the Input and the Learner.
The learner can be as detailed as the input data needs to analyze it. Among the options to analyze information it features the decision tree analysis, forward selection and backward elimination, weighting optimizing, genetic algorithm, wrapper validation, simple cross validation, and evolutionary feature construction.

It features a data wizard by selecting the input data and the learner to show the result s of the process. Besides it displays a list of available database drivers; in case there were some missing or not available just copy to the drivers to the lib/jdbc directory and restart RapidMiner in order to make them available.

Its installation process is simple and easy to perform; no additional libraries or programs are needed. Documentation is included with the program; it supplies a tutorial and a manual likewise there are much more documentation available on its website.

ii Reliability
Its GUI is easy and friendly to user. Its data wizard guides users step by step during all the process. It supplies options to show the results in graphic mode.

iii Availability
RapidMiner runs in any platform because is open source. But there are binary distributions ready for Windows, Linux and Mac OS platforms. The link provided to download the binary distribution is for Windows platform.

iv Maintainability
The core of RapidMiner is open source. In upcoming platform releases, RapidMiner Radoop will be broadened to support those
security measures early-on that evolve and have the potential to be adopted as security standards within enterprises. With that, RapidMiner Radoop is future-proof delivering easy-to-use in-Hadoop analytics on any Hadoop cluster – no matter what security implementations will be involved

v. Security
Beyond authentication, RapidMiner Radoop now also supports data access authorization employing Apache Sentry. In several distributions, Apache Sentry is used to control access e.g. to tables in Hive.
As with any other configuration requirement, configuration of Kerberos authentication support in RapidMiner Radoop is easy and frictionless. RapidMiner Radoop hides all administration and configuration complexity and reveals only necessary settings to the user. Effectively, configuration and administration requirements for IT concerning RapidMiner Radoop as in-Hadoop analytics solution are reduced to a minimum.
With perimeter security and data access security supported for most Hadoop clusters (given the broad adoption of Kerberos and Sentry), RapidMiner Radoop already delivers security for a large portion of production clusters deployed within organizations.

III. Database Requirements

Initially it was a tool designed to support the Freebase database and community for data cleaning, reconciliation and upload. However, I have not utilized any database for this project. The data set was present in csv format, imported to OpenRefine and re-exported to csv.

IV. Design Specifications

a. Work Done

i. Chosen tool set: OpenRefine

Strengths

1.OpenRefine is a desktop application. It opens in the browser as a Local Webserver. So, the data is safe and it doesn't get uploaded to the Google server.
2.It has facets which is used to filter the data into subsets and these clusters can be customized and organised into meaningful data.
3.It has a Browser based interface, and so can handle more data efficiently.
4.Openrefine has a strong feature in extending data -user can use it to find Meta Data and it can be used to correlate with it.

Weakness

1.The UI of Openrefine is not user friendly. Although the features and functions are strong, the UI make Openrefine looks boring. Besides, in the visualization, the function is not scalable. For instance, Openrefine give user a view of data, but the image is not big enough to figure out complex distribution.
2.Unfortunately Google has removed support for this tool, making few of its features redundant.

RapidMiner

i Strengths

1. Powerful due to its learning operators and operator framework, which allows to form nearly arbitrary processes
2. Results are showed also in graphic mode.
3. It supports and can accept new data drivers.
4. It's easy and friendly to use.
5. It provides wizards for data management.

Weaknesses

1. The system kept crashing on some models when the dataset was large.

ii Correlation

WheelTypeID and WheelType are different representations of each other. PurchaseDate and VehicleAge are closely correlated. On a correlation filter of 0.90, 4 attributes areremoved. On a correlation filter of 0.01, 18 attributes were removed.

iii Classification Model

The classification model used for the final dataset was decision tree model. For the cleaning of WheelType and AUCGUART decision tree is used. For the cleaning of PRIMEUNIT Bayesian Model is used.

VI. Test Documentation

RapidMiner

a. Features to be tested

Testing against test data, testing against training data, testing with split(validation)
Testing against test data
When testing against training data, I got a prediction accuracy of 99.21%.
When testing against testing data, the following results were obtained for each of the split

Training	Test	Accuracy
50%	50%	87.90%
70%	30%	88.04%
71%	29%	88.07%
90%	10%	87.94%

b. Test Cases

1. Purpose

To test accuracy of the model under different conditions

2. Sample Input

{1.0 0 ADESA 2005.0 5.0 CHEVROLET TRAILBLAZER 2WD 6C LS 4D SUV 4.2L MAROON AUTO 65343.0 AMERICAN MEDIUM SUV GM 6194.0 9872.0 6599.0 10679.0 13302.0 52117.0 45005.0 OH 8005.0 0 1020.0 Alloy NO GREEN}

3. Expected Output

{1.0 0 ADESA 2005.0 5.0 CHEVROLET TRAILBLAZER 2WD 6C LS 4D SUV 4.2L MAROON AUTO 65343.0 AMERICAN MEDIUM SUV GM 6194.0 9872.0 6599.0 10679.0 13302.0 52117.0 45005.0 OH 8005.0 0 1020.0 Alloy NO GREEN}

4. Test Procedure

The validation operator takes as input an ExampleSet. Based on the parameters provided, it splits the data into two parts. The tuples belonging to each part are chosen at random. The classification model is trained using the first part and the values of the second part is predicted using the second part. This is used to compute the values of precision, recall, recognition and accuracy.