Research Journal of Applied Sciences, Engineering and Technology 10(10): 1108-1132, 2015
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: January 31, 2015 Accepted: March 12, 2015 Published: August 05, 2015

Research Article

A Study of Open Source Data Mining Tools and its Applications

P. Subathra, R. Deepika, K. Yamini, P. Arunprasad and Shriram K. Vasudevan
Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham (University), Coimbatore, India

Abstract: Data Mining is a technology that is used for the process of analyzing and summarizing useful information from different perspectives of data. The importance of choosing data mining software tools for the developing applications using mining algorithms has led to the analysis of the commercially available open source data mining tools. The study discusses about the following likeness of the data mining tools-KNIME, WEKA, ORANGE, R Tool and Rapid Miner. The Historical Development and state-of-art; (i) The Applications supported, (ii) Data mining algorithms are supported by each tool, (iii) The pre-requisites and the procedures to install a tool, (iv) The input file format supported by each tool. To extract useful information from these data effectively and efficiently, data mining tools are used. The availability of many open source data mining tools, there is an increasing challenge in deciding upon the apace-updated tools for a given application. This study has provided a brief study about the open source knowledge discovery tools with their installation process, algorithms support and input file formats support.

Keywords: Data mining, KNIME, ORANGE, Rapid Miner, R-tool, WEKA

INTRODUCTION

In the present era, the value and the explosive rate of data in the database and data repository storage seems to be increasing which has made the humans impossible to manually analyze them for valuable decision-making. This is the phase where Data Mining comes into picture. The various Data Mining techniques makes it easier for the users for effective data extraction (Hand *et al*., 2001) (i.e.,) transformation of vast amount of data into useful strategic information and provides knowledge that is of a great use in the field of marketing, business, banking, science and engineering, games, bioinformatics etc.

Thus, Data Mining plays an active part in the discovery of strategic patterns which is used to find relationship among the data using various data analytics tools by applying various statistical and mathematical concepts (Rangra and Bansal, 2014). At this period of time, there is an measurable amount of data mining tools available in the market such as, KNIME ("About Knime", https://www.knime.org/), R-Tool (Spector, 2004), WEKA (Hall *et al*., 2009) ORANGE and Rapid Miner etc., each of it has its set of own methods and algorithms that will be useful for effective and efficient extraction and utilization of meaningful data from the Data Warehouse (storage of historical data) (Wahbeh *et al*., 2011) and thus making it easy for the users in the knowledge gaining process.

In this research we have made a comparative study of 5 major data mining tools depending on their individual features, compatibility and degree of correctness in the classification of the data along with its accuracy. Thus, this comparative study will make things easier for a beginner to understand the usage of the data mining tools and to decide upon which tool can be used for the process accordingly. A basic evaluation of these various tools are done and their installation procedures, supported algorithms for clustering and classification processes are been tabulated along with their supported input file formats.

MATERIALS AND METHODS

Historical development and state-of-art:

KNIME: Konstanz Information Miner (KNIME) is one of the open source data mining tool. This acme software was developed at the University of Konstanz headed by Michael Berthold from Silicon Valley in January 2014.

Features: KNIME ["About Knime", https://knime.org] is purely based on the nodal work which incorporates more than 100 nodes for analyzing the data. The modular approach of KNIME helps in documenting and storing the order in which the analysis processes are conceived and implemented ensuring that the intermediate results are always available to the user. KNIME achieves scalability through sophisticated data handling (catching the data automatically in the background while increasing the throughput performance). We can also import/export workflows in order to exchange the work modules with other KNIME

Corresponding Author: P. Subathra, Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham (University), Coimbatore, India

users. It can also incorporate WEKA analysis modules and R scripts through additional plug-ins. It supports Parallel execution on multi-core systems and Command line version.

Versions: The entry open source version of KNIME is the KNIME Desktop which provides limited features. Paid versions of KNIME is available with more of additional features and with great supporting factors. It is based on the Eclipse IDE which makes it as a development platform similar to data mining platform. The following table shows the available versions of KNIME tool and its supported platforms.

R tool: R is best known as a statistical tool for analyzing data. It is a simple programming language based on S programming language. It was initially created by Ross Ilaka and Robert gentleman and was recently developed by *R* development core team at the University of Auckland, New Zealand.

Feartures: R is an interpreted language (Paradis, 2002) (effective programming language) that is, in most computer languages like C, FORTRAN and Pascal requires built option in executing a command but R differs from them by executing commands directly. R provides an easy programming environment for its users and includes options for data summary, data exploration, graphical presentation and data modeling. It can also incorporate other R functions that can often be incredibly useful for analyzing and visualizing data. The facilities like data manipulation, calculation, graphical facilities and display are provided by this software. It contains some intermediate tasks flexibility to the users.

WEKA: Waikato Environment for Knowledge Analysis (WEKA) is popular machine learning toolkit developed at the University of Waikato, New Zealand and is mainly use in academia (Hand *et al*., 2001).

Features: Waikato Environment for Knowledge Analysis is developed at the University of Waikato (Hall *et al*., 2009), New Zealand. It provides toolkits for machine learning. WEKA provides a comprehensive collection of algorithms. WEKA provides itself with the GUI, so algorithms can be easily used to the dataset from GUI directly or it can be called from the Java code. It also supports working in the command prompt. Tasks like preprocessing, feature selection, clustering, can be done using WEKA. It is written in Java so it is platform independent and can run in almost any platform. It also supports visualization tasks and many machine learning applications. WEKA is freeware available under the General public license agreement (GNU).

Pre-requirements: The following table shows Java version necessary to run a specific WEKA version.

ORANGE: ORANGE ("Orange-Datamining", http://orange.biolab.si/) is a component-based suite which provides a programming front-end for visualization and analysis of data. Its user interface builds upon Qt framework which is a cross-platform. Orange is distributed free under the GPL. It uses python libraries to do scripting works. Shortest script for doing training, cross validation. The easiest tool to learn. It is implemented in C++ and Python. The invention of this software was done at University of Ljubljana, Slovenia.

Features: Tasks like data preprocessing, feature scoring and filtering are done using this software. Unlike its competitor scikit-learn and mlpy, Orange does not tie into NumPy and its ecosystem of tools; it focuses on traditional, symbolic algorithms, more than numeric ones. It also supports model evaluation and exploration techniques, algorithms comparison and prediction. Works both as a script and interface.

Versions:
Rapid miner: Rapid Miner, formally known as YALE ("Rapid Miner-Open Source Datamining", http://it.toolbox.com/wiki/index.php/RapidMiner) (Yet another Learning Environment) was developed by Ralf Klinkenberg, Ingo Mierswa Simon Fischer in 2001 in the technical University of Dort mung is freeware suite. It is one of the comprehensive, flexible and the most widespread-used due to the combination of its leading-edge technologies and its functional range.

Rapid Miner can be used as an intuitive Graphical User Interface (GUI) or as a command line version for extracting patterns. Like other software processes like Visualization, optimization, modeling, construction is supported.

Software versions: Community edition (open-source) Enterprise edition (Community Edition + More Features + Services + Guarantees).

APPLICATIONS SUPPORTED

KNIME:
Customer intelligence: KNIME ("Knime Use cases", http://www.knime.org/applications,) is used in the Telecommunication industry for the identification of classes of telecommunication customers with the help of K-Means clustering methods. It allows the users to give the different number of clusters for the K-Means to calculate since the number of clusters is unknown beforehand.

Finance: Credit scoring-KNIME workflow deals with the creation of credit scoring model which based on the historical data that uses Decision tree, neural networks and SVM at the same time to specify which model is more accurate for different modeling techniques so that it can be validated to ensure the accuracy rate. The output of the workflow is in the PMML format so that it can be used for different applications.

Manufacturing:
Energy usage prediction: This KNIME workflow is used to predict the future values of the time series with respect to the past values. This workflow helps in removing the seasonality from the time series and also trains an auto-regressive model for time series prediction.

Pharma:
Virtual high throughput screening: KNIME is designed to handle huge data sets so the results of predictive analysis of the untested data can be easily visualized with the help of the enrichment plotter.

Chemical library enumeration: This KNIME workflow helps the chemists to create a virtual library of amides based on the amines and acids. Further some of the molecular properties of the enumerated products are calculated and filtered based upon the Lipinski "rule of 5". This workflow uses RDKit, CDK and indigo integration nodes to demonstrate interoperability.

Retail:
Social media music recommendation: We can use KNIME's predictive analysis algorithm to recommend music preferences for the in the social media. It also helps in transforming the data to make it more suitable for applying association and advanced association algorithms to list the artists and recommendations of interest.

Cross industry: Address duplication-Typos are common mistakes of any user, but restaurant names, postal services and many other services depends on the correct string pattern. Therefore string matching is an important part of data analytics. KNIME uses string matching algorithm to get rid of typo errors and finds out the closest string to what the user wrote. By taking the string distance between our current misspelled data and the reference data set.

Government: Network traffic reporting: This workflow of KNIME is used for generating a report of the connection statistics of the IP addresses of the hosts in a network. Dataset is obtained by capturing a package of data using WIRESHOCK and it is converted in CSV file format. The CSV file is used KNIME for further analysis to generate the report on connection statistics. This workflow of KNIME is available under its Server package.

Cell miner: KNIME is used for analyzing cell images. A new data cell is integrated for images and picture file reader node is added to the repository. A segmenter node is also used to identify the nodes in the images. Multiple feature extraction nodes are also used to extract the data in order to input it to the classifier algorithm.

R tool:
Data exploration and visualization: R includes all kinds of data manipulation, statistical model which is needed for Data Analysis technique with various merits ("R Language Features", http://www.revolutionanalytics.com/what-r). It supports Unique Data Visualization for representing the complex data into charts and graphs. R provides flexibility to mix-and-match the models which provides best performance (Ramamohan *et al*., 2012).

Outlier detection: The Extreme values package of R provides outlier detection and plot functions for univariate data. The parameters for a model distribution are estimated using regression of the sorted data obtained from the subset of data on their QQ-plot positions.

Text mining: The tm package of R provides the text mining framework which has methods for data import, corpus handling, preprocessing, metadata management and creation of term-document matrices.

R is also used for Association rules mining, Social network analysis, Multimedia scaling, Parallel computing etc.

WEKA:
Machine learning: WEKA is used in solving a variety of real-world problems like agricultural and horticultural domains.

Data visualization: The 3D visualization perspective is a plug-in Spoon perspective for WEKA version 4.0 and above provides a Java based 3D scatter plot visualization and a histogram matrix.

Time series and analysis: WEKA has a dedicated time series analysis framework that allows forecasting models to be developed, evaluated and visualized. This environment helps in modeling time series by transforming the data into a form that can be processed by the standard propositional learning algorithms.

WEKA is also used in the field of Text mining and Fraud detection.

ORANGE:
Bioinformatics: ("Orange Add-ons", http://orange.biolab.si/) Bioinformatics add-on provides a way for analytics that is useful in gene selection, quality control, scoring distances between experiments. It also provides public access to PIPAx database, GEO data sets, Biomart, GO, Atlas, Array Express, KEGG. These features can be combined with powerful visualization, network exploration and data mining techniques from the ORANGE data mining framework.

Model maps: ORANGE Model Maps add-on extends Orange by providing modules for building model maps and also provides widgets for Orange Canvas that enables exploration of model maps by the users.

Network analysis: ORANGE Network add-on that the users with provides network visualization and analysis tools.

Text mining: ORANGE Text Mining add-on extends Orange by providing common functionality for basic text mining tasks that makes ease for the users.

Rapid miner:
Different levels of analysis those are available: Artificial neural networks-Non-linear predictive models that resemble biological neural networks in structure.

Genetic algorithms: Optimization techniques that use processes such as genetic combination, mutation and natural selection in a design based on the concepts of natural evolution.

Other applications:

- Automatic abstraction
- Financial forecasting
- Targeted marketing
- Medical diagnosis
- Credit card fraud detection
- Weather forecasting

INPUT DATA FORMATS

The following table shows the input data formats that are accepted by each tool under study:

- Java Sript Option Notation (.json)
- Extensible Markup language (.xml)
- Attribute Relation file format (.arff)
- Comma separated values (.csv)
- Atom
- Binary file format (.bin)
- C4.5 file format (.c45)
- Tab-separated values file Format (.TSV)
- .xrff – extended attribute relation file format

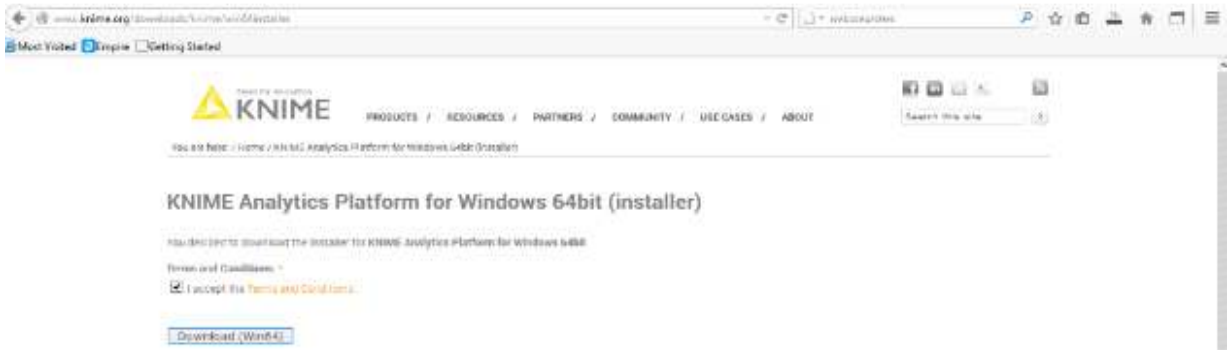


Fig. 1: KNIME download webpage

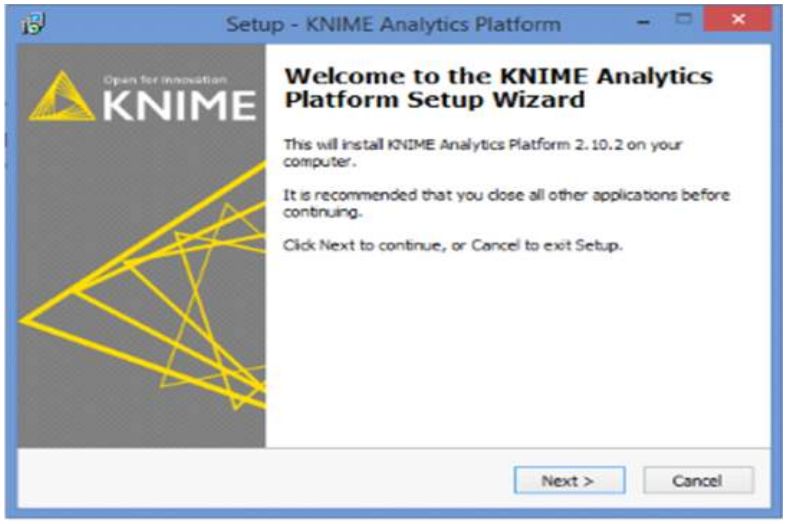


Fig. 2: Setup run process

INSTALLATION STEPS

KNIME:

Instructions: Download the package according to your Windows version and system type.

Step1 : Download the KNIME Software Package from http://www.knime.Org/ (Fig. 1).

Step 2: Run the setup.exe file and follow the below instructions (Fig. 2).
Click next to view the license agreement.

Step 3: Carefully read the license agreement and accede the agreement as above (Fig. 3).

Step 4: Give the installation path location for KNIME Analytics Platform (Fig. 4).

Step 5: Create the path location for shortcut files of KNIME Analytics Platform and click next (Fig. 5).

Step 6: Click the check box icon in order to create a desktop icon and click next (Fig. 6).

Step 7: Click install to proceed (Fig. 7).

Wait until the installation process is over. It may take 4-6 min to get installed (Fig. 8 and 9).

Now the KNIME Analytics Package will be successfully installed.

R tool:

Step 1: Select the language to use during the installation and click Ok button (Fig. 10).

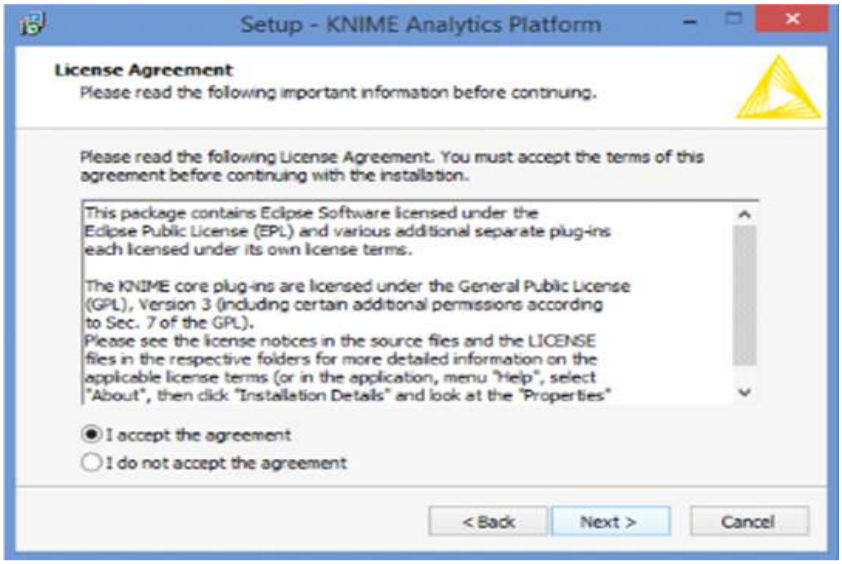


Fig. 3: License information

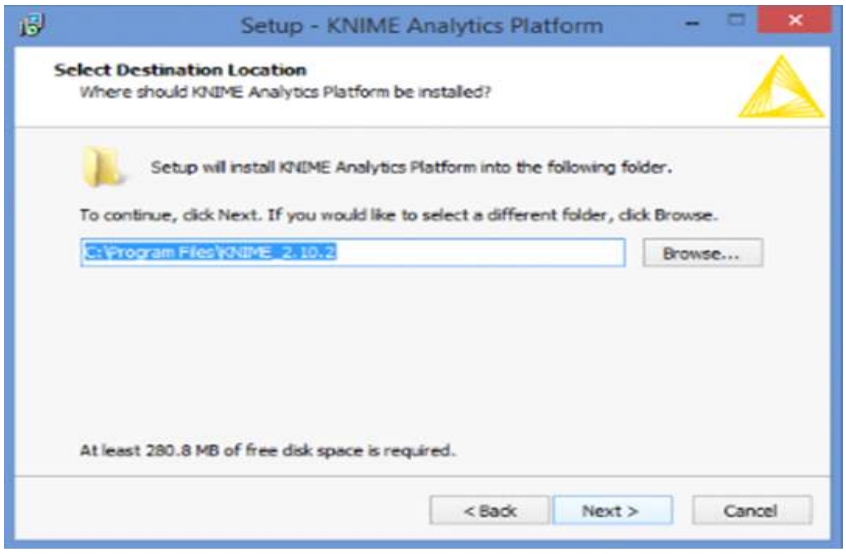


Fig. 4: Installation guidelines

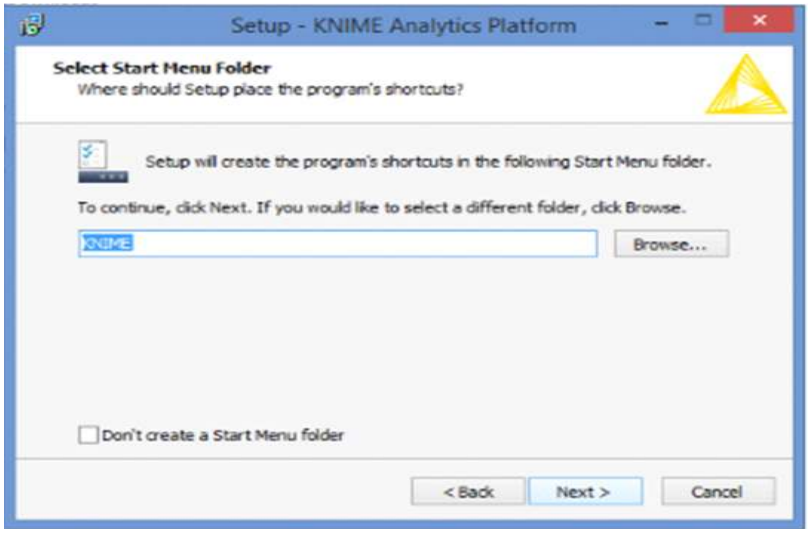


Fig. 5: Path information for installation

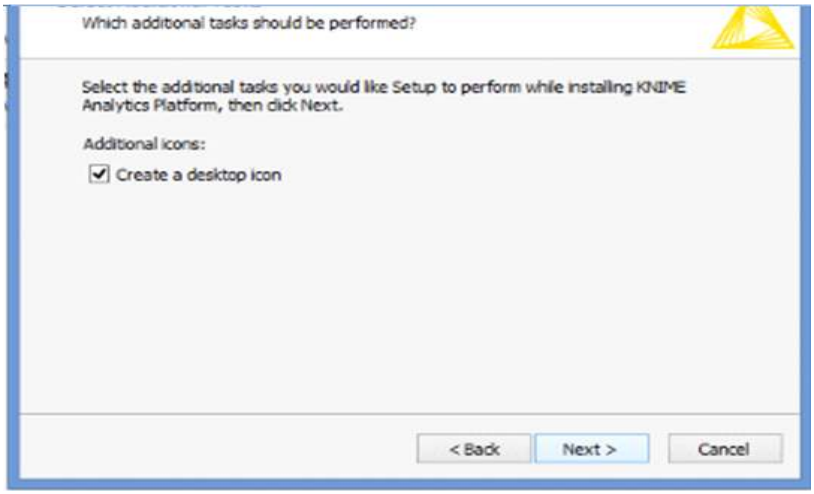


Fig. 6: Procedure for creation of desktop ikon

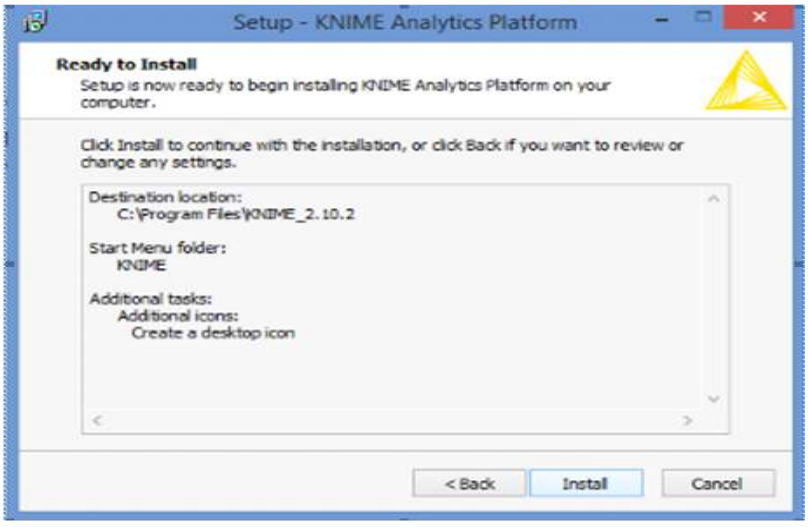


Fig. 7: Final step in installation

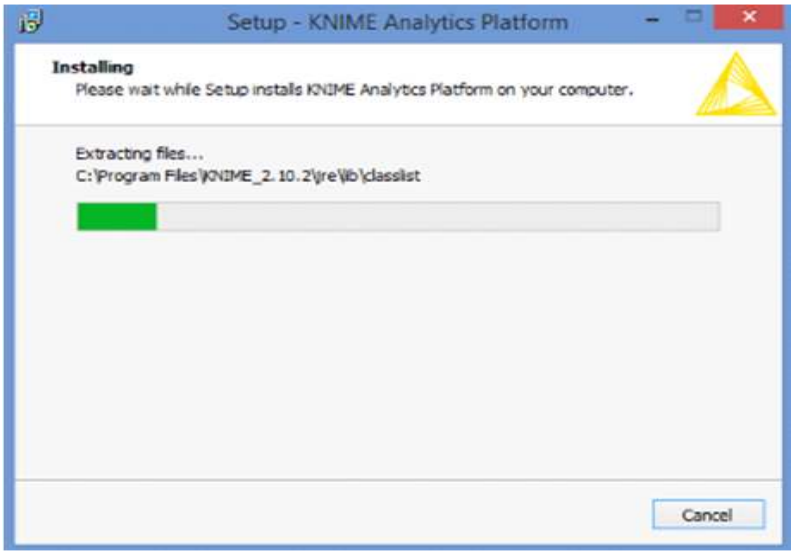


Fig. 8: Installation progress

Fig. 9: Finished installation

Fig. 10: Language selection for R tool

Step 2: Click next and proceed (Fig. 11).
Step 3: Read the terms and conditions to proceed (Fig. 12).
Step 4: Select the destination location to save the installation (Fig. 13).
Step 5: Select the appropriate system file and proceed further (Fig. 14).
Step 6: Set the startup configuration if required (Fig. 15).
Step 7: Select the path location to place the program shortcuts (Fig. 16).
Step 8: Choose the available options to customize else Check all the check boxes for default settings and click next to proceed (Fig. 17).

Fig. 11: Guidelines to proceed

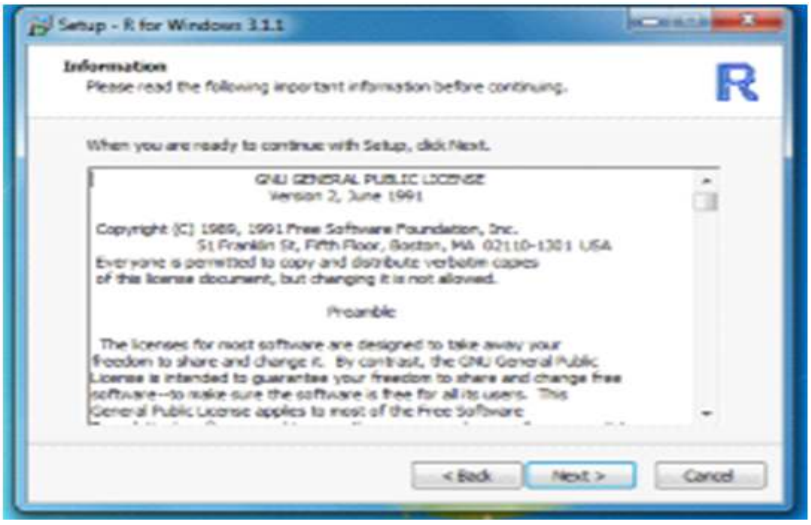


Fig. 12: License information

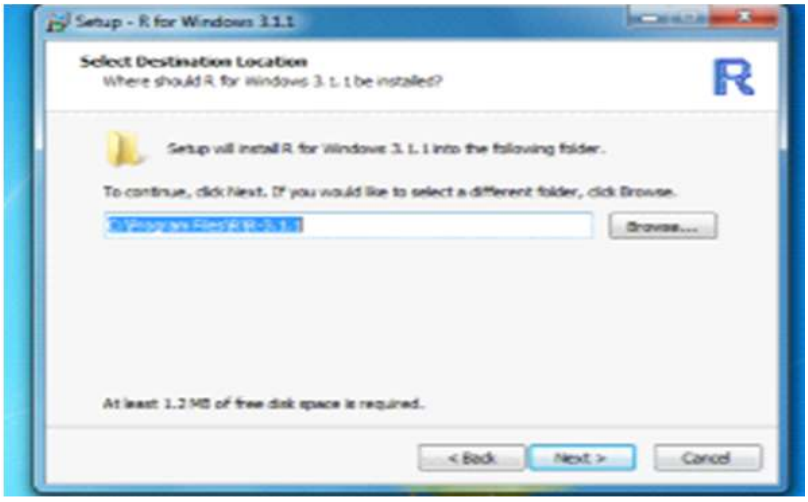


Fig. 13: Installation guidelines

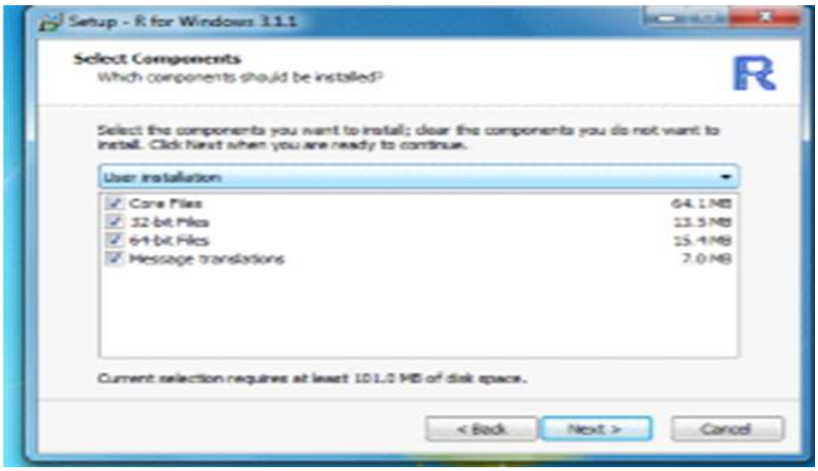


Fig. 14: System file selection process

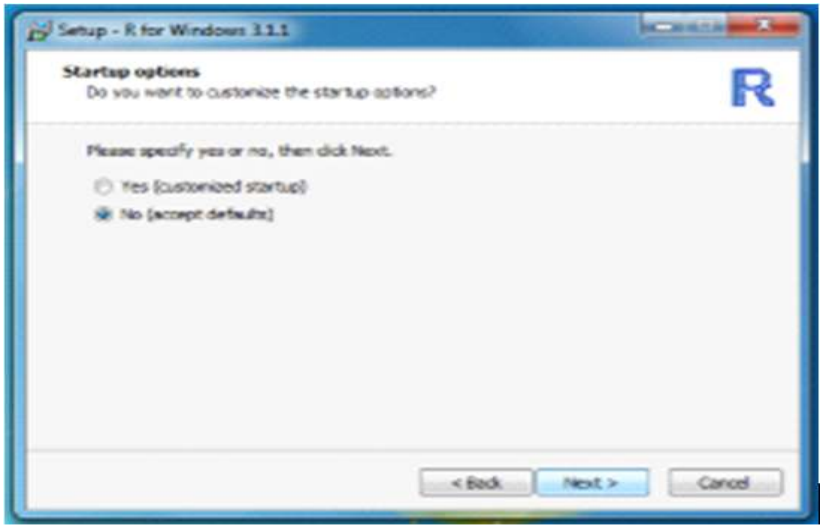


Fig. 15: Startup configuration

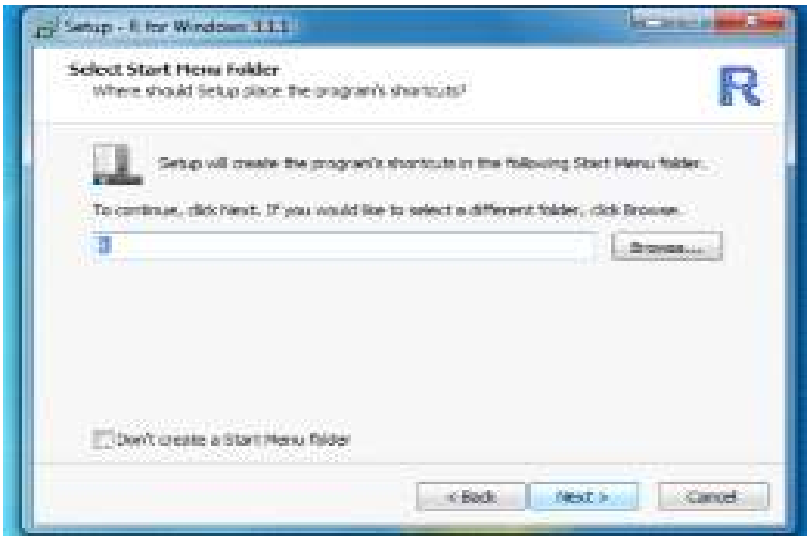


Fig. 16: Path information for installation

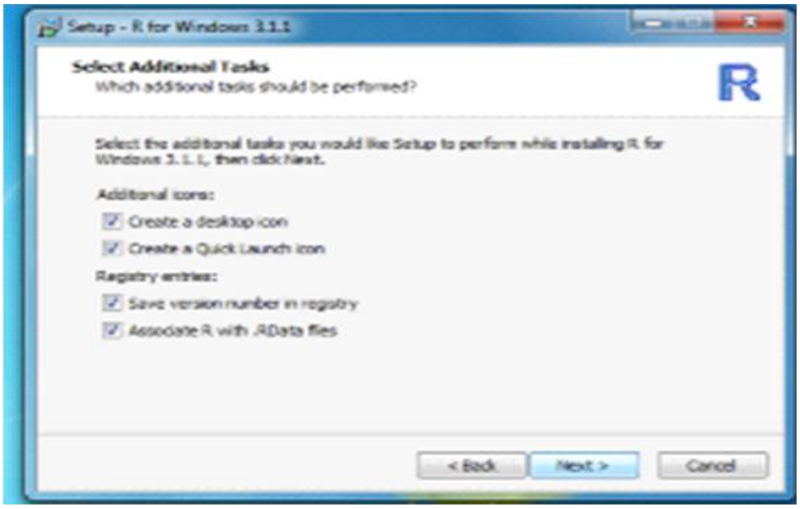


Fig. 17: Path location information

Fig. 18: Settings configuration

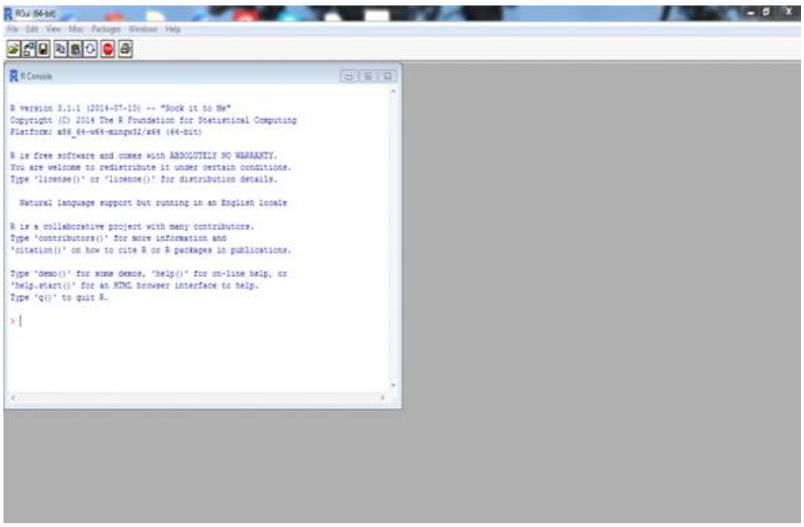


Fig. 19: Installation finish step

Step 9: Click finish (Fig. 18).
Package will successfully installed (Fig. 19).

WEKA:
Step1 : Run setup.exe (Fig. 20).
Click next button to continue.
Step 2: Read the instructions carefully and accede it (Fig. 21).
Step 3: Check all the packages and click next button to continue (Fig. 22).

Fig. 20: Setup run process

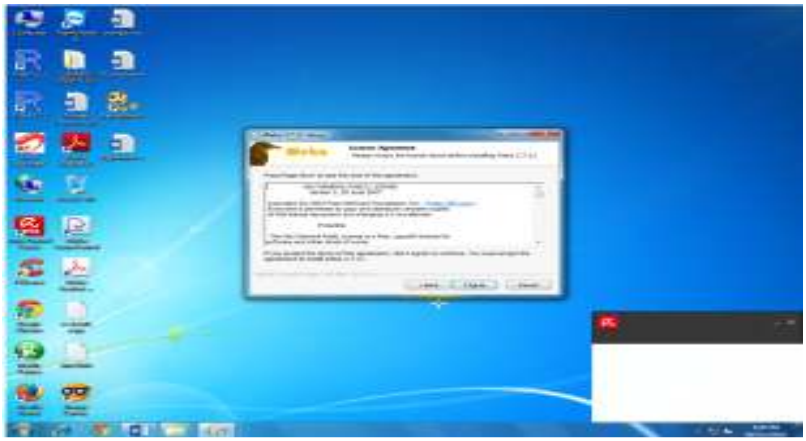


Fig. 21: License information

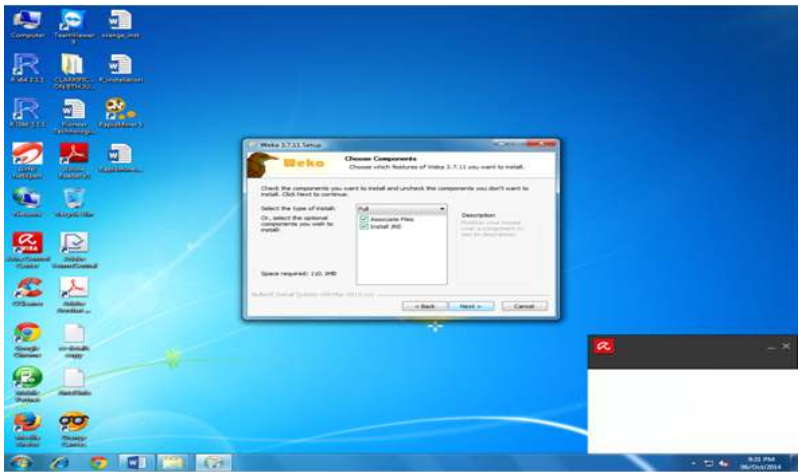


Fig. 22: Package selection

Step 4: Select the installation path and click next button to continue (Fig. 23).

Step 5: Give a name for the shortcut and specify the folder in which the shortcut should appear (Fig. 24).

Step 6: Install java if it is not already installed in your system. Click install button and continue (Fig. 25).

Step 7: Click close button and proceed (Fig. 26).

Step 8: Click next and continue (Fig. 27).

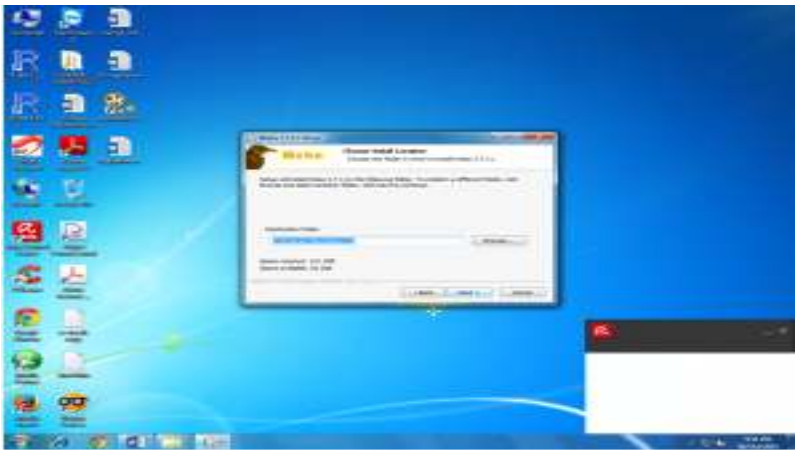


Fig. 23: Path information for installation

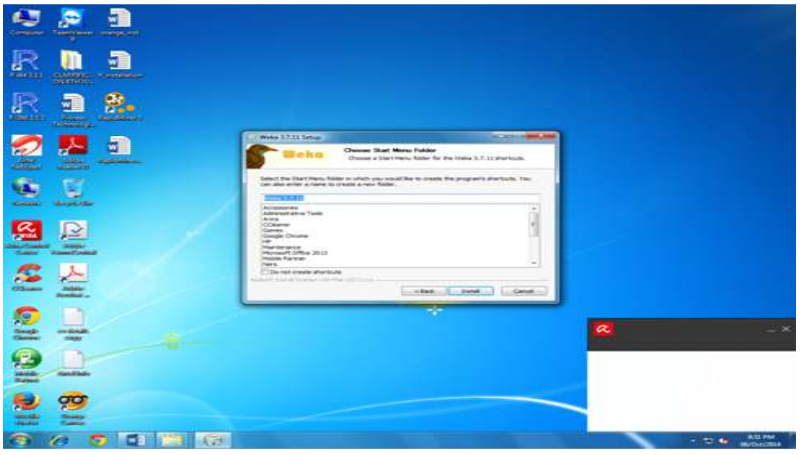
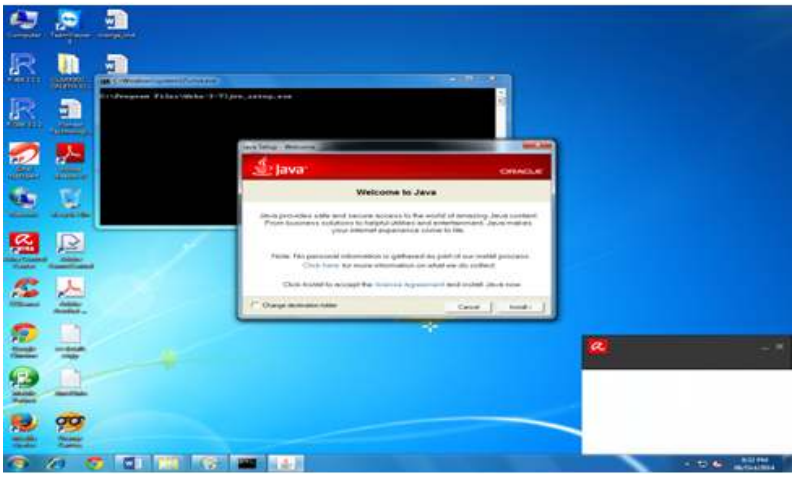


Fig. 24: Path location information



(a)

(b)

Fig. 25: Java installation

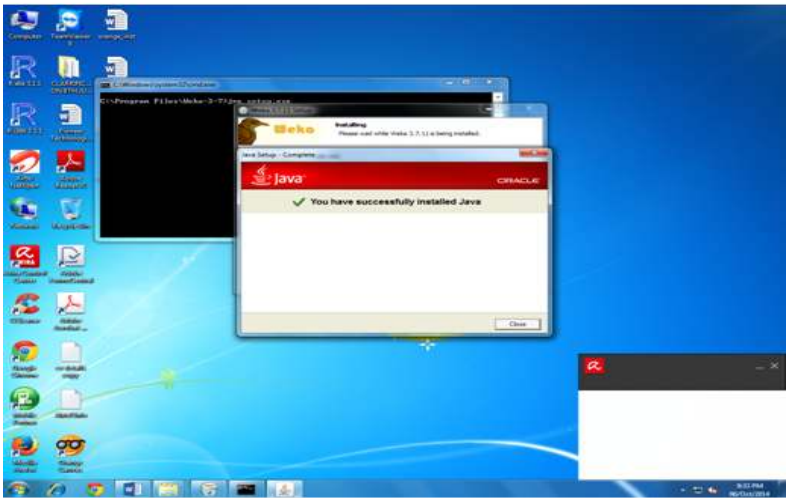


Fig. 26: Installation progress enablement

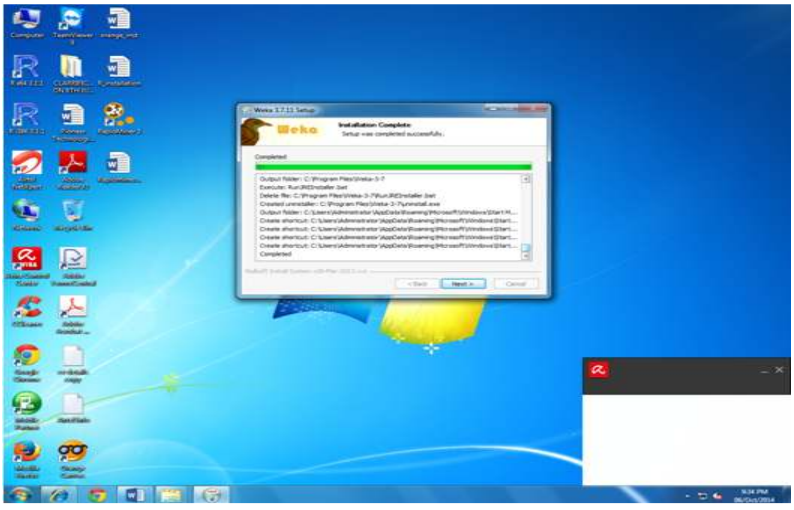


Fig. 27: Settings configuration

Step 9: Check Start WEKA and click on finish button (Fig. 28).
Now the WEKA will be successfully installed (Fig. 29).

ORANGE:

Step 1: Run setup.exe. Read the instructions carefully and accede it (Fig. 30).

Step 2: Click yes button and continue (Fig. 31).

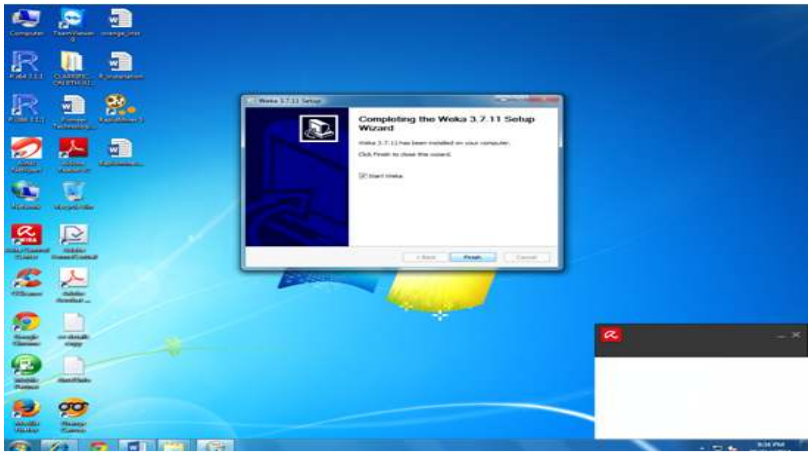


Fig. 28: Installation finish step

Fig. 29: Successful installation message

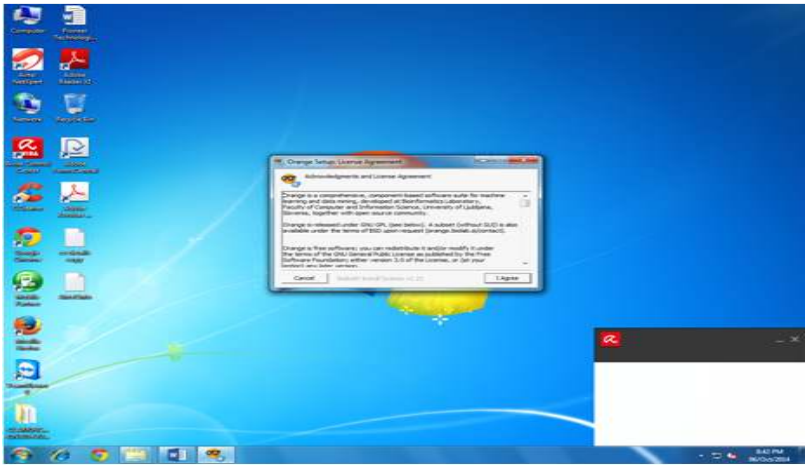


Fig. 30: Setup run process

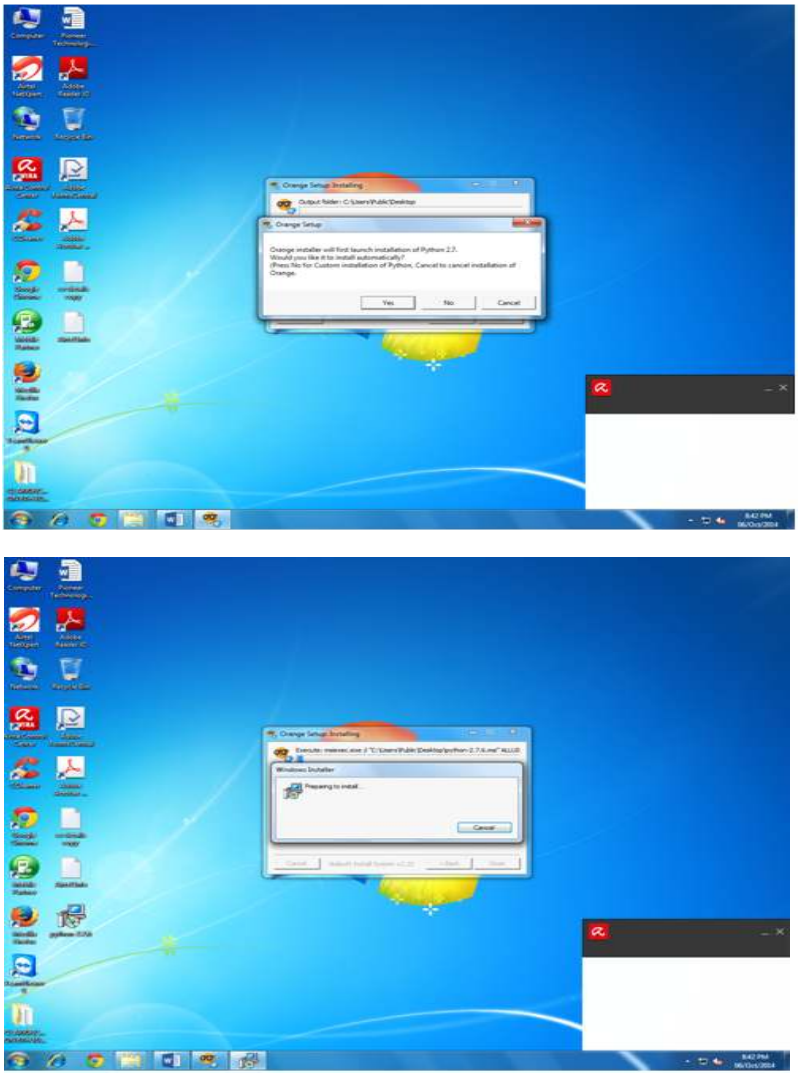


Fig. 31: Acceptance to accelerate the installation

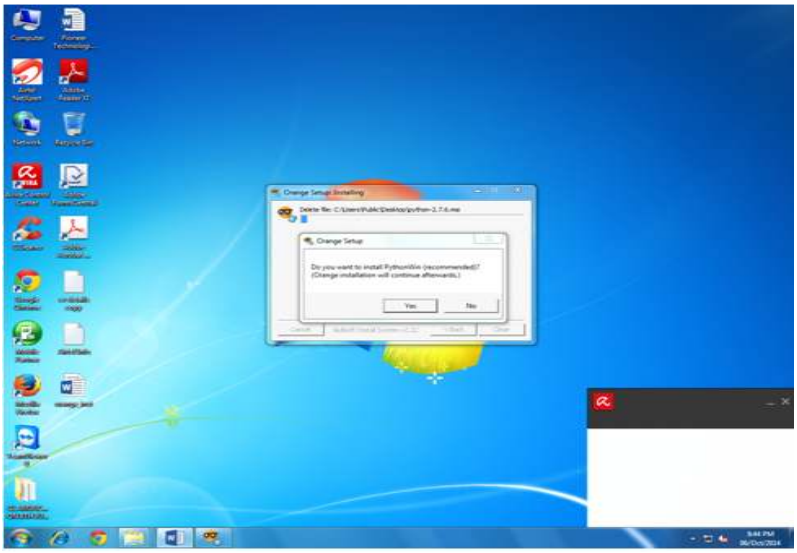


Fig. 32: Python installation

Step 3: To install python Click yes button and continue (Fig. 32).
Step 4: Click next button and continue (Fig. 33).
Step 5: Click next button and continue (Fig. 34).
Step 6: Click next button and continue (Fig. 35).
Step 7: PyQt setup wizard appears. Click next button and continue (Fig. 36).
Step 8: Read the license agreement carefully and click on 'I Agree' button to proceed (Fig. 37).
Step 9: Select the Desired components to be installed. For default settings select check all the check boxes (Fig. 38).
Step 10: Select the installation path and click next button to continue (Fig. 39).

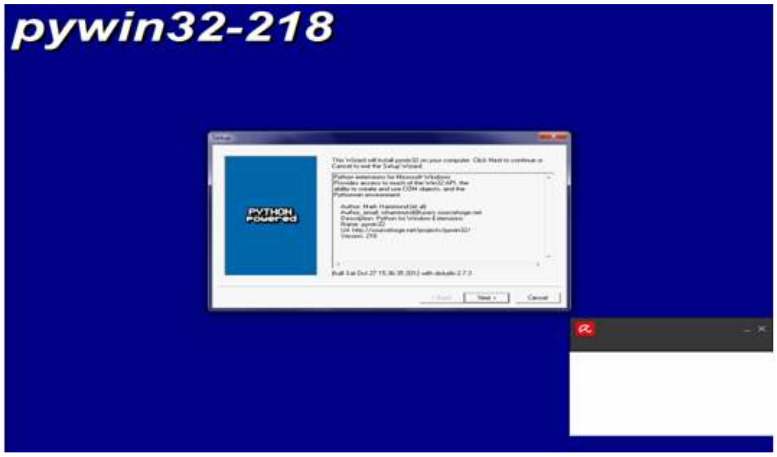


Fig. 33: Installation process

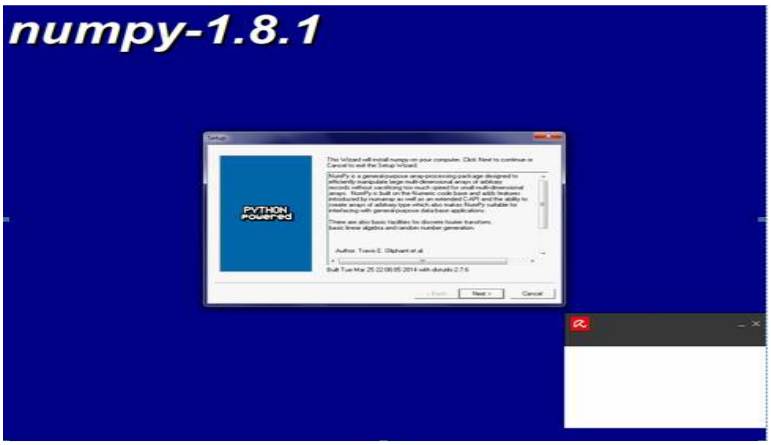


Fig. 34: Installation process

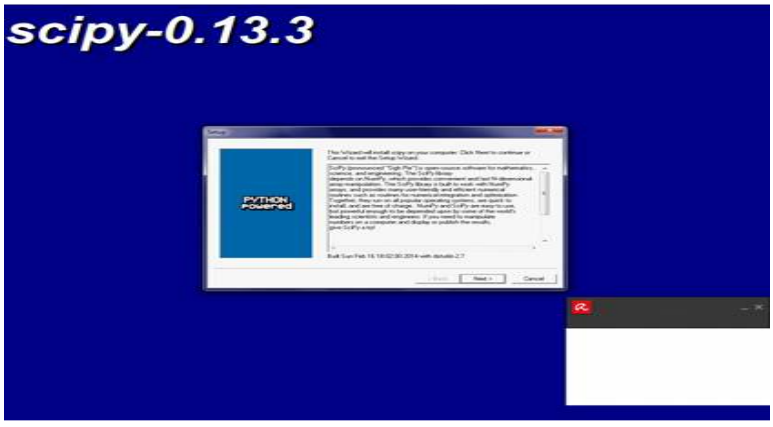


Fig. 35: Installation process

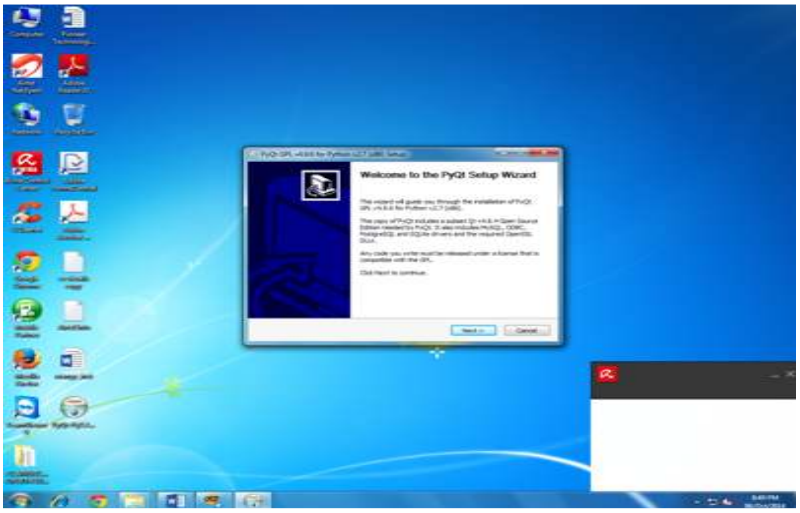


Fig. 36: Confirmation with PyQt setup wizard

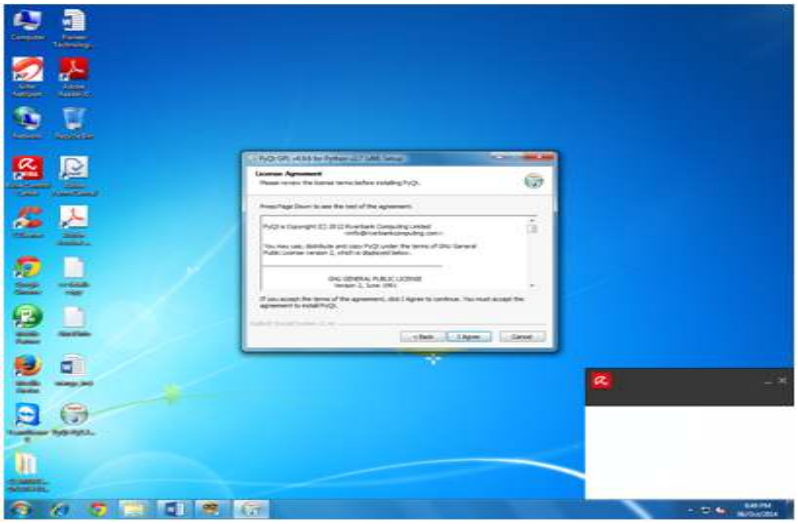


Fig. 37: License information

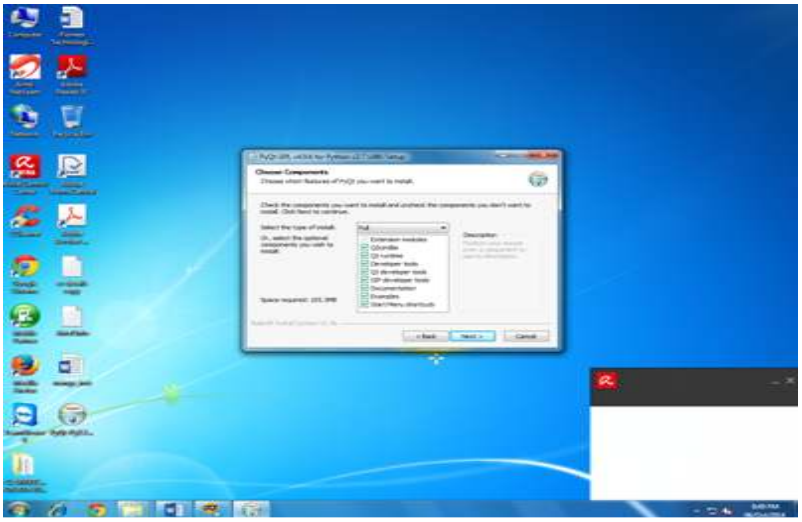


Fig. 38: Customized configuration

Step 11: Click finish button (Fig. 40). The PyQt package will be successfully.

Step 12: PyQwt setup wizard appears. Click next button and continue (Fig. 41).

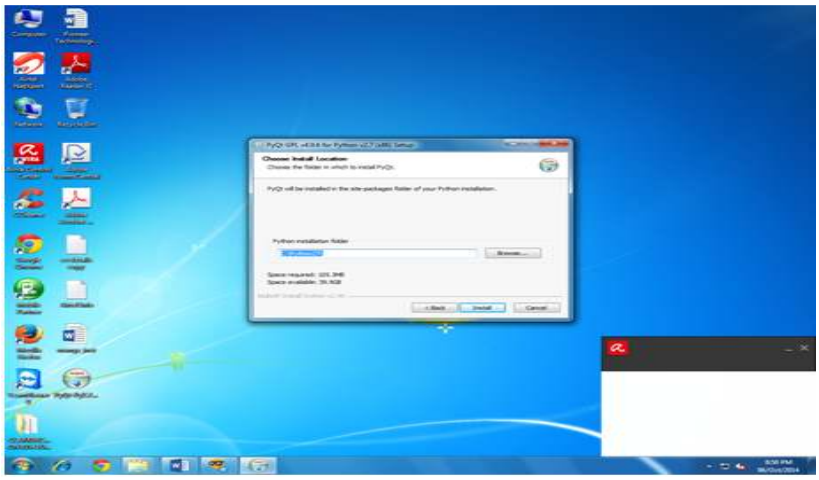


Fig. 39: Progress confirmation

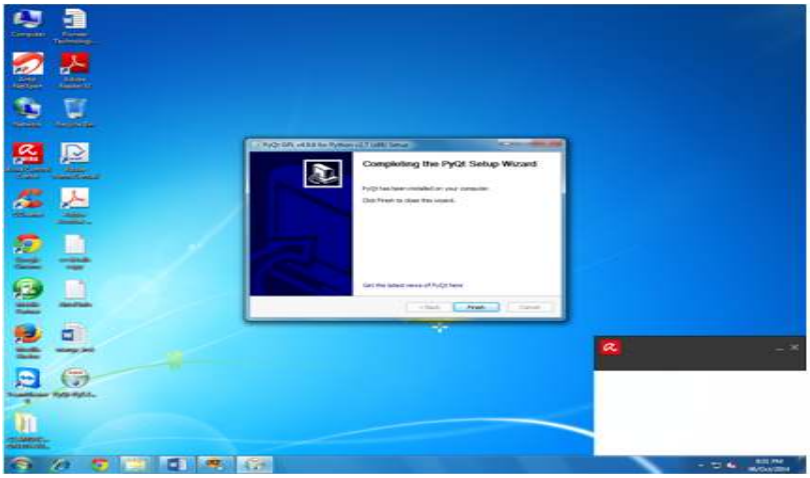


Fig. 40: Installation finish step

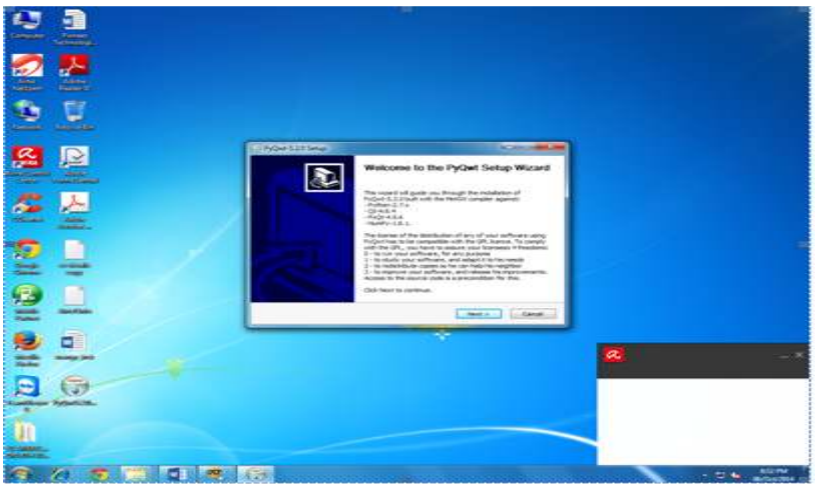


Fig. 41: PyQwt setup wizard window

Step 13: Read the license agreement carefully and click on 'I Agree' button to proceed (Fig. 42).

Step 14: Select the Desired components to be installed. For default settings select check all the check boxes (Fig. 43).

Step 15: Select the installation path and click next button to continue (Fig. 44).

Step 16: Click finish button (Fig. 45).
The PyQwt package will be successfully.

Step 17 : Click next button and continue (Fig. 46).

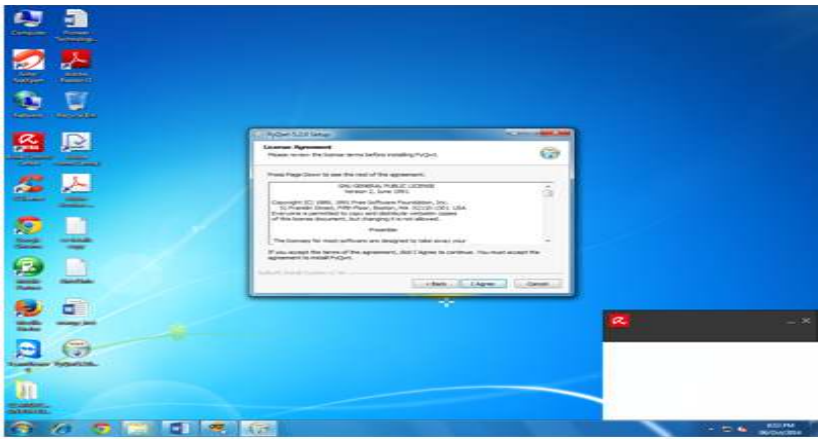


Fig. 42: License agreement

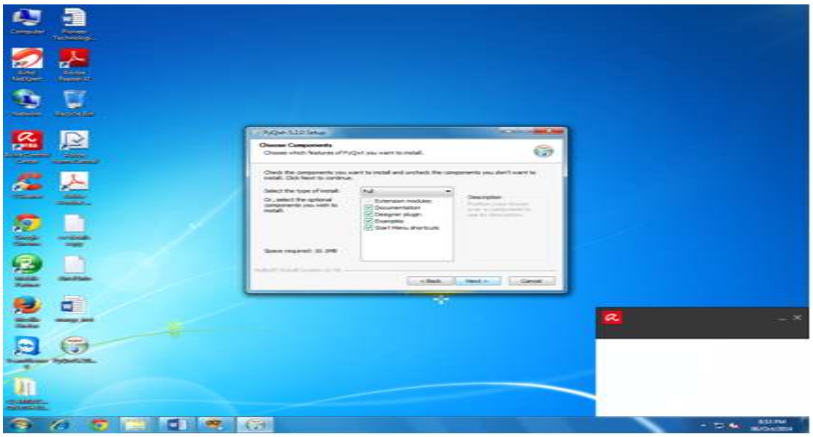


Fig. 43: Customized configuration

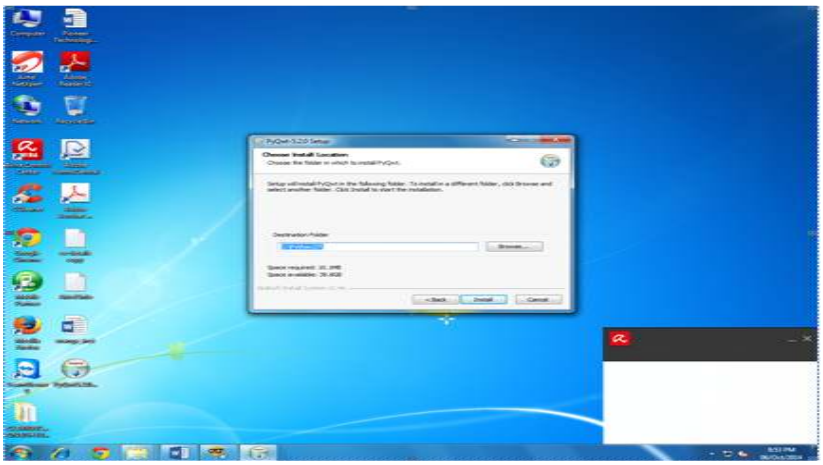


Fig. 44: Path information for installation

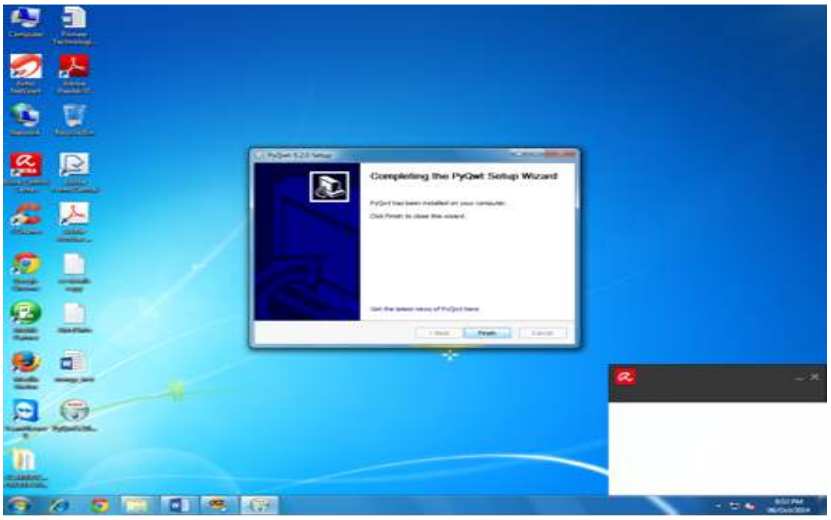


Fig. 45: Installation finish step

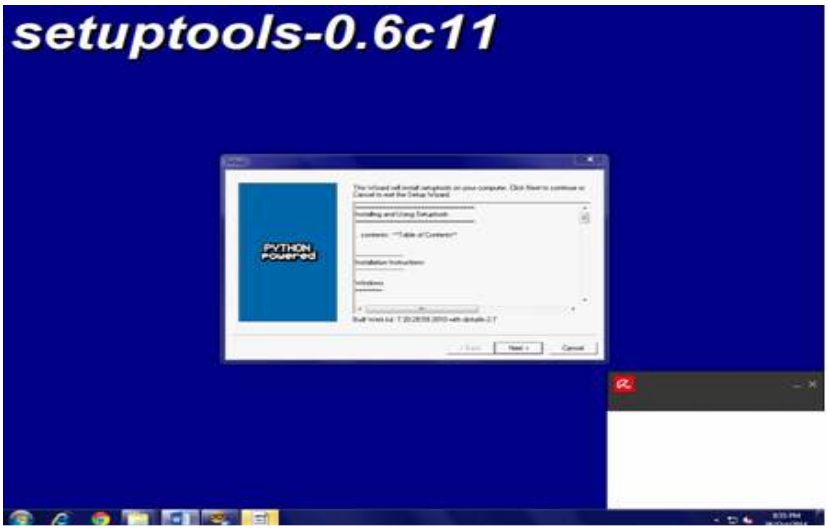


Fig. 46: Click next to move forward

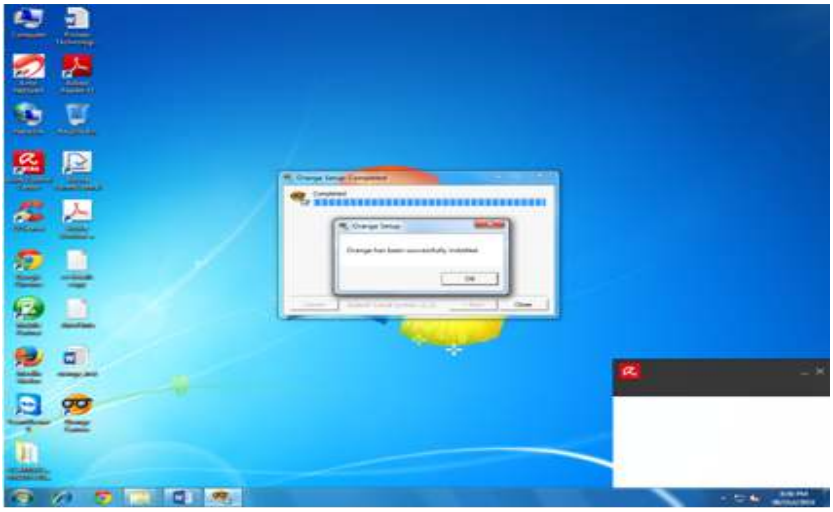


Fig. 47: Final step in installation of ORANGE

Step 18 : Click finish Button (Fig. 47).

Orange tool will be successfully installed in your system (Fig. 48).

Rapid miner:

Step 1: Run Setup.exe file to start installation (Fig. 49).

Step 2: Read the instructions carefully and accede it (Fig. 50).

Step 3: Choose the installation path for the package (Fig. 51).

Step 4: Installation complete (Fig. 52).

Step 5: Click Finish button to complete installation (Fig. 53).

Fig. 48: Installation begins

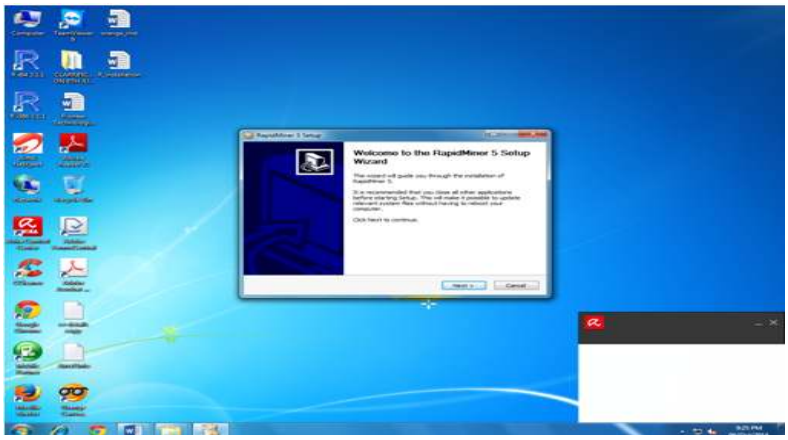


Fig. 49: Instructions on license

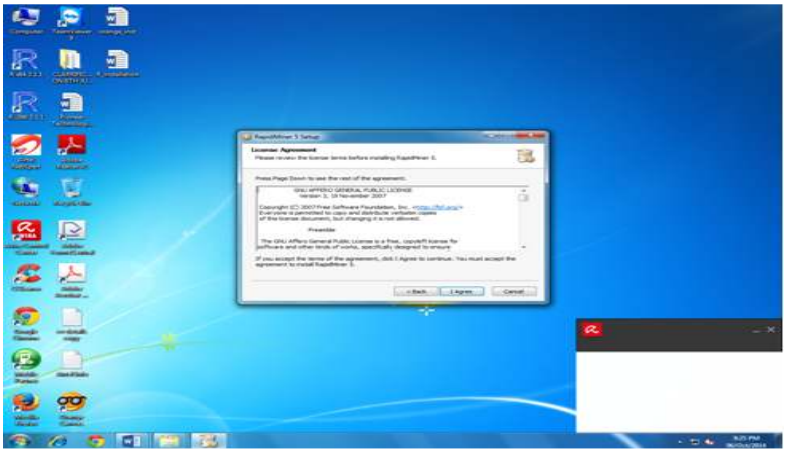


Fig. 50: Path selection

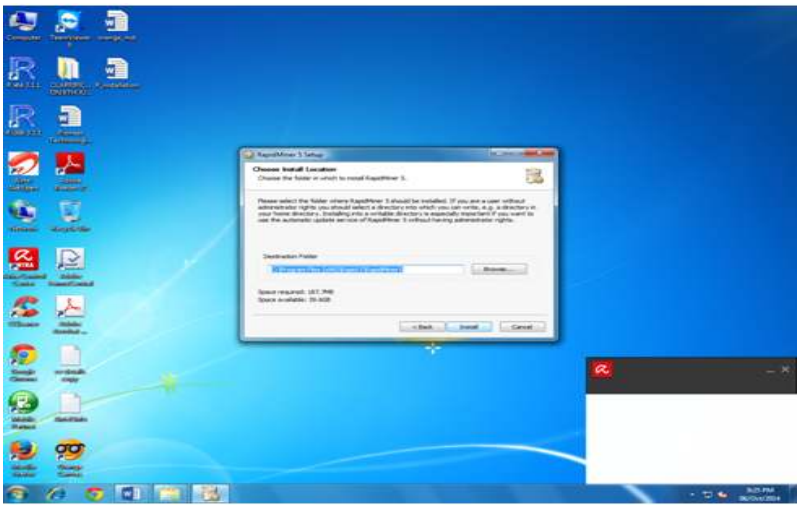


Fig. 51: Installation completion

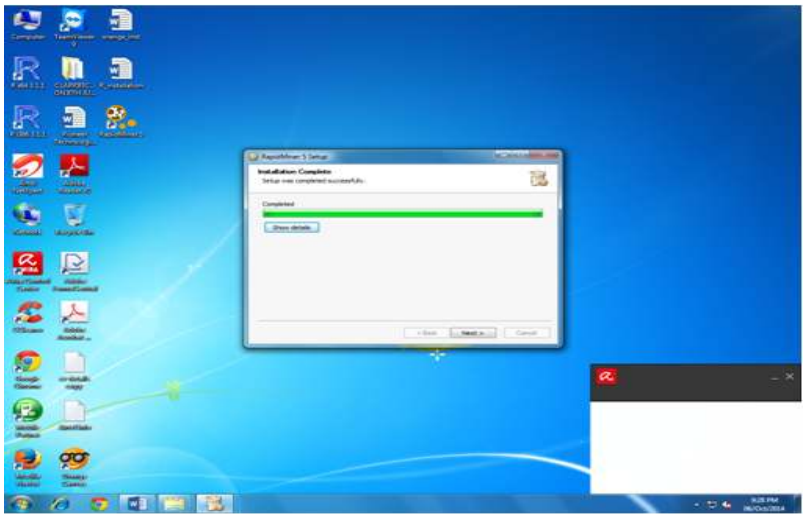


Fig. 52: Final step (finish)

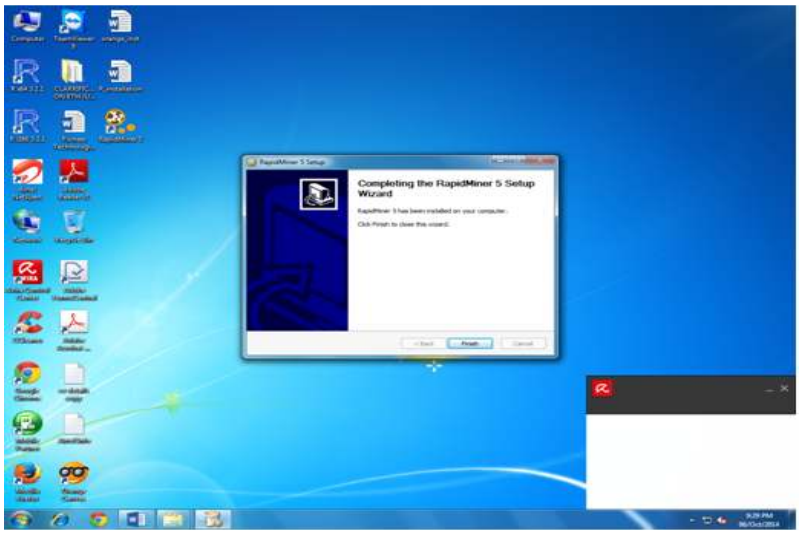


Fig. 53: Final step (finish)

Table 1: KNIME available versions and platforms

Version	System type (bit)	Available platform
KNIME	32	Linux, Windows
	64	Linux, Windows, MacOS X
KNIME developer version	32	Linux, Windows
	64	Linux, Windows, MacOS X

Table 2: R tool available versions and platforms

	R tool versions
Windows (32/64 bit)	R 2.9.x
	2.10.x
	2.11.x
	2.12.x
MAC OS	2.13.x
UNIX platforms	2.14.x
	R 2.15.x
	R 3.0.x
	R 3.1.2-pumpkin helmet (stable)

Table 3: WEKA version and the supported Java version

	Java version supported			
WEKA version	1.4	1.5	1.6	1.7
<3.4.0	X	X	X	X
3.4.X	X	X	X	X
3.5.X	✓ (3.5.0-3.5.2)	✓ (>3.5.2)	X	X
3.6.X	✓	X	X	X
3.7.X	✓	✓ (3.7.0)	✓ (>3.7.0)	X

A workaround for the problem of look 'n' feel due to the combination of Java 5.0 and later versions with the Gnome/Linux, was introduced with version 3.4.5/3.5.0. From WEKA m 3.6.5/3.7.4 Mac OS X users will need Java package version updated

Table 4: WEKA available platform

WEKA, WEKA developer version	Available platform	Executable file available
	Window X86	• With java VM 1.7 • Without java VM
	Window X64	• With java VM 1.7 • Without java VM
	MAC OS X	• Oracle's java VM 1.7 • Apple's java VM 1.6
	Linux	• Without java

If you already have Java 1.6 (or later) on your system then go for the version without java VM

Table 5: ORANGE versions and platforms available

Platforms	Versions
Windows,	Orange 2.5
Mac OS X, Linux	Orange 2.6
	Orange 2.7
Add-ons	Versions available
Orange-bioinformatics	2.5.37
Orangecontrib.earth	0.1.3
Orange-textable	1.4.2
Orange-text	1.2a1
Orange- reliability	0.2.14
Orange-NMF	0.1.2
Orange-multitarget	0.9.3

ALGORITHM SUPPORT

Clustering algorithms: Clustering is a process of partitioning a set of abstract data/objects into classes of similar objects by assigning the labels to the groups. These sub-classes are known a clusters. The following table shows the various clustering algorithms supported by the data mining tools.

Classification algorithm: In finding and retrieving data in effective and efficient manner, Data classification plays a vital role. It organizes data into categories. The following table shows the classification algorithms that are supported by data mining tools.

RESULTS AND DISCUSSION

Thus, from the Table 1 to 9 one can understand the basic components of the 5 major Data Mining tools which has been discussed above along with its acceptable file formats and the various algorithms that can be used for clustering and classification processes.

Table 6: Rapid miner versions and available platforms

Version 6.2 (latest)	Windows (32 and 64 bit) Macintosh (Mac OS 10.7.3) Linux Requirement:- JAVA 7 should be pre-installed
Version 6.1 (stable)	Windows (32 and 64 bit) Macintosh (Mac OS 10.7.3) Linux Requirement: JAVA 7 should be pre-installed
Version 5.3 (stable)	Windows Macintosh (Mac OS 10.7.3) Linux Requirement: i. JAVA 7 should be pre-installed JAVA 6 should be pre-installed if full server package to be downloaded

	ORANGE	KNIME	Rapid Miner	WEKA	R tool
Platform	Different versions of Linux, Apple's Mac OSX and Microsoft windows	Linux, OSX, Windows	Cross-platform	Cross-platform	Cross-platform
Written in	Python C++, C	JAVA	My SQL	JAVA	

Table 7: Input file format supported by KNIME, R tool, WEKA, ORANGE and Rapid Miner

	Supported file formats			
KNIME	R	WEKA	ORANGE	Rapid Miner
Json, xml, arff, csv, atom	.R native formats (incl. R date and RDS), csv, excel	arff, csv, bin, C4.5	csv, tsv, C4.5	arff, att, csv, xrff

Table 8: Classification algorithms supported by KNIME, R tool, WEKA, ORANGE and Rapid Miner

	Rapid Miner	WEKA	R tool	KNIME	ORANGE
K-means	✓	✓	✓	✓	✓
x-means	✓	✓			
DBSCAN	✓				
Expectation maximization cluster	✓		✓		
Support vector clustering	✓				
Random clustering	✓				
Agglomerative clustering	✓				
Top down clustering	✓				
Flatten clustering	✓				
Extract cluster prototypes	✓				
Cobweb		✓			
Farthest first		✓			
Filtered clusterer		✓			
Hierarchial clusterer		✓		✓	✓
OPTICS		✓			
Sib		✓			
Density-based clustering algorithm		✓		✓	

Table 9: Clustering algorithms supported by KNIME, R tool, WEKA, ORANGE and Rapid Miner

Tool name	Naïve Bayes	Id3	Cart	Chaid	J4.8	C4.5	KNN
Rapid Miner	✓	✓	✓	✓	X	✓	✓
WEKA	✓	✓	X	X	✓	X	✓(IBK)
ORANGE	✓	✓	✓			✓	✓
KNIME		✓	✓	✓			✓
R tool	✓	✓	✓	✓	✓	✓	✓

CONCLUSION

This study has given a brief introduction on the five different data mining tools and their applications-KNIME, ORANGE, R Tool, WEKA and Rapid Miner. Each tool has its own pros and cons. Be that as it may, KNIME, ORANGE, R Tool, WEKA and Rapid Miner have most of the desired characteristics and functions for a fully-functional Data Mining platform and thereby these tools can be used for most of the Data Mining tasks. As a future work, we are going to study and compare the performance of various data mining classification and clustering algorithms for various data mining tools.

REFERENCES

Hall, M., E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and I.H. Witten, 2009. The WEKA data mining software: An update. ACM SIGKDD Explor. Newslett., 11(1): 10-18.

Hand, D.J., H. Mannila and P. Smyth, 2001. Principles of Data Mining. MIT Press, Cambridge, MA. Retrieved from: http://www.cs.waikato.ac.Nzml/ weka/requirements.html- referenece; http:// www.cs.waikato.ac.nz/ml/weka/.

Paradis, E., 2002. R for Beginners. Retrieved from: cran.r-project.org/doc/ contrib/ Paradis-rdebuts_en.pdf.

Ramamohan, Y., K. Vasantharao, C.K. Chakravarti and A.S.K. Ratnam, 2012. A study of data mining tools in knowledge discovery process. Int. J. Soft Comput. Eng., 2(3), ISSN: 2231-2307.

Rangra, K. and K.L. Bansal, 2014. Comparative study of data mining tools. Int. J. Adv. Res. Comput. Sci. Softw. Eng., 4(6), ISSN: 2277 128X.

Spector, P., 2004. Introduction to R. University of California, Berkeley.

Wahbeh, A.H., Q.A. Al-Radaideh, M.N. Al-Kabi and E.M. Al-Shawakfa, 2011. A comparison study between data mining tools over some classification methods. Int. J. Adv. Comput. Sci. Appl., pp: 18-26.