			
Subject name	Business Intelligence Lab		Subject Code: ITL602
Class	TE IT	Semester – VI (CBCGS)	Academic year: 2018-19
Name of Student	Kazi Jawwad A Rahim		QUIZ Score : 03 / 06
Roll No	27	Assignment/Experiment No:	09
Title:	Detailed study of BI tool – Rapid Miner (paper Assignment)		

1. Course objectives applicable: LO3
2. Course outcomes applicable: LO3
3. Learning Objectives: <ol style="list-style-type: none"> To study the uses of Rapid Miner (Open source tool) To study the data mining operations
4. Practical applications of the assignment/experiment: Data mining
5. Prerequisites: <ol style="list-style-type: none"> Basic data mining concepts. Use of open source tools
6. Hardware Requirements: <ol style="list-style-type: none"> PC with minimum 2 GB RAM
7. Software Requirements: <ol style="list-style-type: none"> Windows / Linux Web Browser

8. Viva Questions (if any): (Online Quiz will be taken separately batch-wise) <ol style="list-style-type: none"> What is Rapid Miner? What are the services provided by Rapid Miner? What are the different features of Rapid Miner? How data mining can be done using Rapid Miner?
--

9. Experiment/Assignment Evaluation:			
Sr. No.	Parameters	Marks obtained	Out of
1	Technical Understanding (Assessment may be done based on Q & A <u>or</u> any other relevant method.)		6
2	Neatness/presentation		2
3	Punctuality		2
Date of performance (DOP)		Total marks obtained	10
Date of checking (DOC)		Signature of teacher	

10. Theory:

RapidMiner is a data science software platform developed by the company of the same name that provides an integrated environment for data preparation, machine learning, deep learning, text mining, and predictive analytics. It is used for business and commercial applications as well as for research, education, training, rapid prototyping, and application development and supports all steps of the machine learning process including data preparation, results visualization, model validation and optimization.

RapidMiner is developed on an open core model. The RapidMiner Studio Free Edition, which is limited to 1 logical processor and 10,000 data rows is available under the AGPL license. Commercial pricing starts at \$2,500 and is available from the developer.

RapidMiner is a software platform for analytics teams that unites data prep, machine learning, and predictive model deployment.

1. RapidMiner Studio

- Visual Workflow Designer

INCREASE PRODUCTIVITY ACROSS THE ENTIRE DATA SCIENCE TEAM, FROM ANALYSTS TO EXPERTS

1. Accelerate the creation of predictive models using a drag + drop visual interface
2. Rich library of over 1500 machine learning algorithms and functions to build the best model for any use case
3. Pre-built templates for common use cases including customer churn, predictive maintenance, fraud detection, and many more
4. Unique Wisdom of Crowds feature provides proactive recommendations at each step of the workflow, including the population of parameters

- Data Sources

CONNECT TO ALL OF YOUR DATA, NO MATTER WHERE IT LIVES

1. Create point + click connections to databases, warehouses, cloud sources, documents, social media, and business applications
2. Connect to new data sources by downloading extensions from the RapidMiner Marketplace

- In-database Processing

RUN DATA PREP AND ETL PROCESSES DIRECTLY INSIDE DATABASES

1. Query and retrieve data without writing complex SQL
2. Harness the power of highly scalable database clusters
3. Supports MySQL, PostgreSQL, and Google BigQuery

- Explore and Visualize Data

EVALUATE DATA HEALTH, COMPLETENESS, AND QUALITY

1. Explore data using robust statistical overviews and over 30 interactive visualizations
2. Develop an understanding of patterns, trends, and distributions in the data with scatter plots, histograms, line charts, parallel coordinates, box plots, and more.
3. Identify and fix common data quality problems including missing values and outliers

- Data Prep and Blending

ELIMINATE THE HASSLE OF PREPARING DATA FOR PREDICTIVE MODELLING

1. Extract, join, filter, and group data across any number of sources
2. Create repeatable data prep and ETL processes that can be scheduled and shared
3. Jump over to RapidMiner Turbo Prep for a fully interactive point + click data prep experience

- Machine Learning

CREATE ROBUST MACHINE LEARNING MODELS WITHOUT WRITING CODE

1. Choose from hundreds of supervised and unsupervised machine learning algorithms
2. Implement a wide variety of ML techniques including regression, clustering, time-series, text analytics, and deep learning
3. Use both automated and manual feature engineering to improve model accuracy
4. Integrates with RapidMiner Auto Model to create models in 5 clicks using automated machine learning

- Model Validation

UNDERSTAND THE TRUE PERFORMANCE OF A MODEL BEFORE DEPLOYING TO PRODUCTION

1. Eliminate overfitting through a unique approach that prevents model training pre-processing data from leaking into the application of the model
2. Add proven techniques, like cross validation, to a model with just a single mouse click

- Explainable Models Not Black Boxes

CREATE VISUAL DATA SCIENCE WORKFLOWS THAT ARE EASY TO EXPLAIN AND EASY TO UNDERSTAND

1. Each step in the data prep, modeling, and validation process is documented for complete transparency

2. Visual workflow is easy to explain to others in the organization
3. Supports the Local Interpretable Model-Agnostic Explanations (LIME) framework

- Flexible Scoring and Deployment

TURN PREDICTIVE INSIGHTS INTO PRESCRIPTIVE ACTIONS

1. Quickly deploy scored data to spreadsheets and data visualization tools
2. Turn models into production web services with RapidMiner Server
3. Add RapidMiner Real-Time Scoring for demanding high transaction / low latency use cases

- Automation and Process Control

BUILD SOPHISTICATED VISUAL WORKFLOWS AND AUTOMATE IMPORTANT TASKS

1. Use process control operators to create workflows that repeat and loop over tasks, branch flows and access system resources
2. Supports a variety of scripting languages for custom integrations and automations
3. Schedule processes

- Open and Extensible

INTEGRATE WITH EXISTING APPLICATIONS AND CODE

1. Use existing R and Python code and libraries to extend RapidMiner
2. Download new functionality through the RapidMiner Marketplace
3. RapidMiner Studio open core is available under an aGPL license

2. RapidMiner Server

Share and re-use predictive models, automate processes, and deploy models into production

- Team Collaboration

ONE PLACE TO SHARE, MANAGE, AND SECURE DATA PREP AND MODELING PROCESSES

1. Share data prep and modeling processes in a central repository
2. Configure granular permissions to control access to specific processes and folders

- Process Automation

AUTOMATE IMPORTANT TASKS AS OFTEN AS NEEDED

1. Create scheduled processes to prep and clean data, retrain models, and continuously score data in real-time
2. Integrate with external applications through a REST API

- Lightning Fast Model Creation

USE DEDICATED SERVER HARDWARE TO RADICALLY SPEED UP PREDICTIVE MODEL CREATION

1. Take full advantage of multi-core multiprocessor server architectures
2. Push jobs from RapidMiner Studio to RapidMiner Server in a single click

- Turn Insight into Action

OPERATIONALIZE PREDICTIVE MODELS AND TURN PRESCRIPTIVE ACTIONS INTO PRESCRIPTIVE RECOMMENDATIONS

1. Create production web service APIs in just a few mouse clicks
2. Deploy models to RapidMiner Real-Time Scoring for high volume, low latency scoring
3. Monitor model performance over time to detect degradation and retrain as needed

- Scalable, Reliable, and Secure

DEPLOY ON A MODERN ARCHITECTURE BUILT FOR MISSION-CRITICAL DATA SCIENCE APPLICATIONS

1. Deploy on-premise or in the cloud using pre-configured images for Microsoft Azure and AWS
2. Run in a highly available active/active configuration to minimize downtime risk
3. Uses modern authentication, authorization and encryption standards

3. RapidMiner Radoop

Eliminate the complexity of data science on Hadoop and Spark

- Code free machine learning for Hadoop & Spark

BUILD AND RUN PREDICTIVE MODELS IN HADOOP WITHOUT HAVING TO CODE IN SPARK.

1. Create predictive models using the RapidMiner Studio visual workflow designer
2. Expand beyond MLlib to tackle a broader set of use cases including time series and text analytics
3. Supports HDP 3, Cloudera 6, and MapR 6

- Harness the power of Hadoop clusters

RUN DATA PREP AND MACHINE LEARNING JOBS DIRECTLY INSIDE HADOOP

1. RapidMiner *SparkRM* enables all operations and data process flows in RapidMiner Studio to run in-parallel inside Hadoop
2. Jobs are automatically translated into Spark and Hive
3. No additional software is required in the Hadoop cluster environment

- Supports Hadoop standards and security

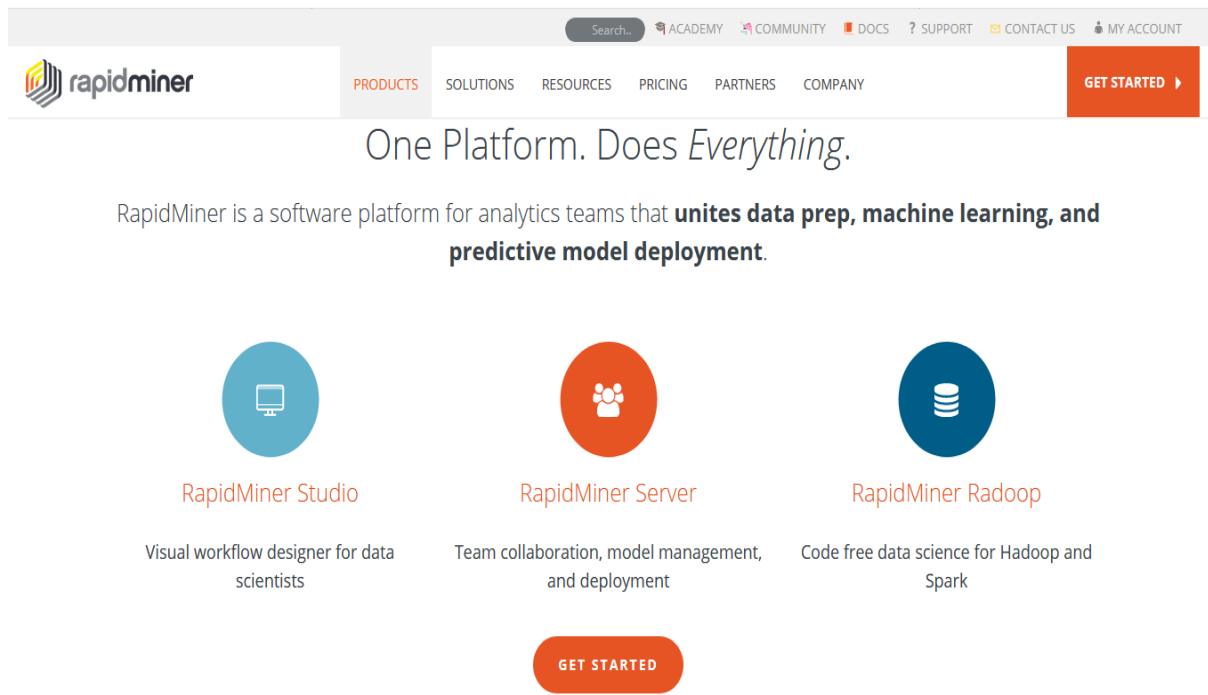
MAXIMIZE YOUR INVESTMENT IN THE HADOOP ECOSYSTEM

1. Re-use existing SparkR, PySpark, Pig, and HiveQL code
2. Reduce risk and enforce regulatory compliance with built-in Apache Sentry & Apache Ranger support
3. Deploy HDFS encryption to comply with data security policies

11. Performance Steps:

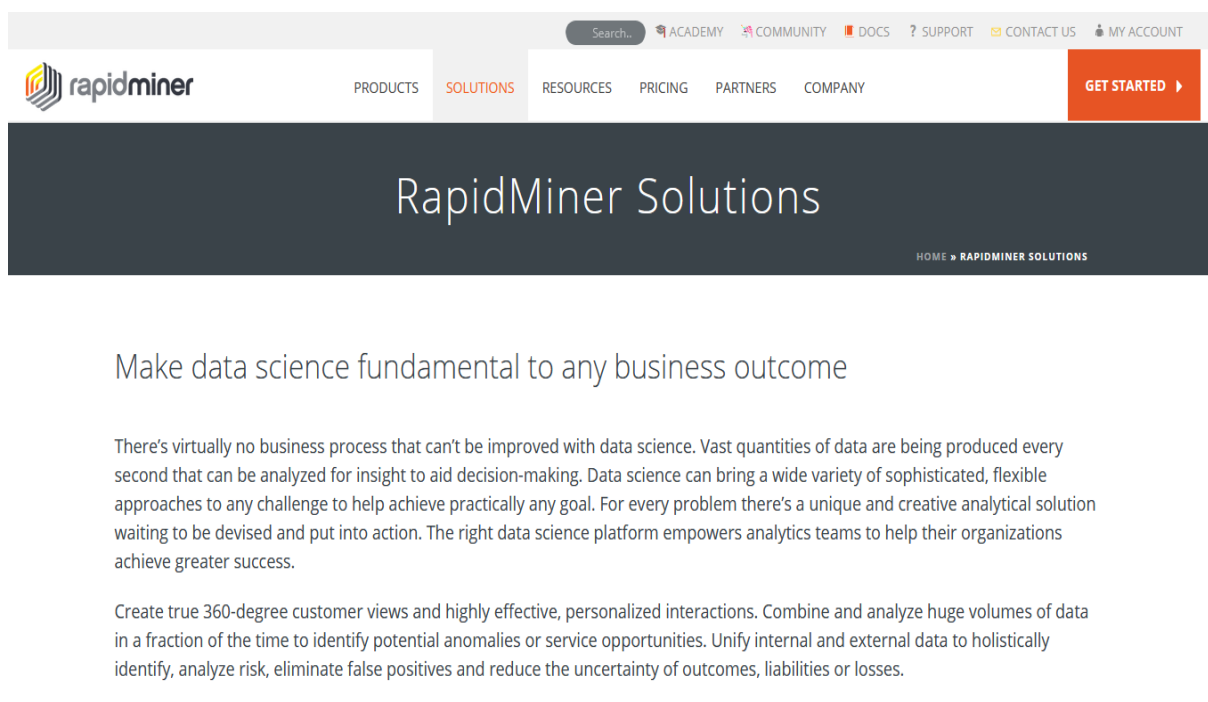
1. Visit the site <https://rapidminer.com/>
2. Search and study the Product and Services.

12. Results:



The screenshot shows the RapidMiner homepage. At the top, there is a navigation bar with a search bar and links for ACADEMY, COMMUNITY, DOCS, SUPPORT, CONTACT US, and MY ACCOUNT. Below this is a secondary navigation bar with links for PRODUCTS, SOLUTIONS, RESOURCES, PRICING, PARTNERS, and COMPANY. The main heading is "One Platform. Does *Everything*." followed by the tagline "RapidMiner is a software platform for analytics teams that **unites data prep, machine learning, and predictive model deployment**." Below this, there are three circular icons representing different products: RapidMiner Studio (blue circle with a computer icon), RapidMiner Server (orange circle with a group of people icon), and RapidMiner Radoop (dark blue circle with a database icon). Each icon is accompanied by a brief description: "Visual workflow designer for data scientists", "Team collaboration, model management, and deployment", and "Code free data science for Hadoop and Spark" respectively. A "GET STARTED" button is located at the bottom center.

Products in RapidMiner



The screenshot shows the "RapidMiner Solutions" page. The navigation bar is similar to the homepage, but the "SOLUTIONS" link is highlighted. The main heading is "RapidMiner Solutions" with a breadcrumb trail "HOME » RAPIDMINER SOLUTIONS". Below the heading, there is a paragraph stating "Make data science fundamental to any business outcome". This is followed by two paragraphs of text explaining the benefits of data science. The first paragraph states: "There's virtually no business process that can't be improved with data science. Vast quantities of data are being produced every second that can be analyzed for insight to aid decision-making. Data science can bring a wide variety of sophisticated, flexible approaches to any challenge to help achieve practically any goal. For every problem there's a unique and creative analytical solution waiting to be devised and put into action. The right data science platform empowers analytics teams to help their organizations achieve greater success." The second paragraph states: "Create true 360-degree customer views and highly effective, personalized interactions. Combine and analyze huge volumes of data in a fraction of the time to identify potential anomalies or service opportunities. Unify internal and external data to holistically identify, analyze risk, eliminate false positives and reduce the uncertainty of outcomes, liabilities or losses."

Solutions in RapidMiner

13. Learning Outcomes Achieved

1. Understanding of RapidMiner tool.
2. Understanding of Services in RapidMiner.
3. Understanding of Products in RapidMiner.

14. Conclusion:

- 1. Application of studied technique in industry:** Data mining.
- 2. Engineering Relevance:** Collaboration with IoT and Cloud Technology.
- 3. Skills Developed:** Understanding of RapidMiner applications.

15. References:

- [1] <https://rapidminer.com/solutions/>
- [2] <https://rapidminer.com/products/>
- [3] <https://en.wikipedia.org/wiki/RapidMiner>