REPUBLIC OF TÜRKİYE

ALTINBAŞ UNIVERSITY

Institute of Graduate Studies

Electrical and Computer Engineering

# ROBOTIC PROCESS AUTOMATION (RPA) FOR THE IMPLEMENTATION ON MACHINE LEARNING BASED RESEARCH

**Noor Jasim Mohammed MOHAMMED**

Master's Thesis

Supervisor

Asst. Prof. Dr. Ayça Kurnaz TÜRKBEN

Istanbul, 2022

# ROBOTIC PROCESS AUTOMATION (RPA) FOR THE IMPLEMENTATION ON MACHINE LEARNING BASED RESEARCH

**Noor Jasim Mohammed MOHAMMED**

Electrical and Computer Engineering

Master's Thesis

ALTINBAŞ UNIVERSITY

2022

The thesis titled ROBOTIC PROCESS AUTOMATION (RPA) FOR THE IMPLEMENTATION ON MACHINE LEARNING BASED RESEARCH prepared by NOOR JASIM MOHAMMED MOHAMMED and submitted on 15/12/2022 has been **accepted unanimously** for the degree of Master of Science in Electrical and Computer Engineering.

Asst. Prof. Dr. Ayça Kurnaz TÜRKBEN
the Supervisor

Thesis Defense Committee Members:

| | | |
|---|---|---|
| Asst. Prof. Dr. Ayça Kurnaz TÜRKBEN | Department of Software Engineering, Altınbaş University | ___________________ |
| Asst. Prof. Dr. Abdullahi Abdu IBRAHIM | Department of Computer Engineering, Altınbaş University | ___________________ |
| Asst. Prof. Dr. Serdar KARGIN | Departmentof Biomedical Engineering, İstanbul Arel University | ___________________ |

I hereby declare that this thesis meets all format and submission requirements of a Master's thesis.

Submission date of the thesis to Institute of Graduate Studies: ___/___/___

I hereby declare that all information/data presented in this graduation project has been obtained in full accordance with academic rules and ethical conduct. I also declare all unoriginal materials and conclusions have been cited in the text and all references mentioned in the Reference List have been cited in the text, and vice versa as required by the abovementioned rules and conduct.

Noor Jasim Mohammed MOHAMMED

Signature

# DEDICATION

I dedicate and promise this research project to my supervisor, who is significant for guiding me during the entire research project, to my family, who has always supported me during difficult times.

# PREFACE

First and foremost, I would like to thank my supervisor Asst. Prof. Ayça kurnaz turkben for guiding and helping me along the way in writing this dissertation. Discussing my progress, problems, and ideas with my supervisor Asst. Prof. Ayça kurnaz TÜRKBEN a couple of times every week helped me tremendously in understanding the logic behind the research. It made me better realize the technical need for this research work.

# ABSTRACT

## ROBOTIC PROCESS AUTOMATION (RPA) FOR THE IMPLEMENTATION ON MACHINE LEARNING BASED RESEARCH

Mohammed, Noor Jasim Mohammed

M.Sc., Electrical and Computer Engineering, Altınbaş University,

Supervisor: Asst. Prof. Ayça Kurnaz TÜRKBEN

Date: 12/2022

Pages: 57

Recently, the concept of robotic process automation has been implemented widely in the area of digital business to improve productivity. However, exploiting this automation concept in the (machine learning) prediction-based research has not been thoroughly investigated and tested. Therefore, this research study the implementation of Robotic Process Automation (RPA) in the case study of predicting cloud User ID through the attributes of their logs and traces.

This is conducted by proposing an approach model of applying the automation principle in mimicking the manual tasks of prediction research tasks using Artificial Neural Networks (ANN). Accordingly, the proposed approach model consists of the research tasks of (data gathering, data pre-processing, ANN implementation, and Result validation).

Then these research tasks are tested manually and through the automation model using automation software (Macro-recorder). The results show that using the proposed model of PRA in conducting prediction research has a significant effect in reducing time-consumption and increasing accuracy.

The results will be of interest to the research community in computer engineering and the automation study of research activities.

**Keywords:** Robotic Process Automation, Artificial Neural Networks, Research Tasks, KNIME.

# TABLE OF CONTENTS

**Pages**

# LIST OF FIGURES

**Pages**

# ABBREVIATIONS

| | | |
|---|---|---|
| RPA | : | Robotic Process Automation |
| ANN | : | Artificial Neural Networks |
| ML | : | Machine Learning |
| AI | : | Artificial Intelligence |
| PNN | : | Probabilistic Neural Network |
| VDIs | : | Virtual Desktop Infrastructures |
| IBM | : | International Business Machines |
| NNs | : | Neural Networks |
| CSV | : | Comma-Separated Values |
| POI | : | Poor Obfuscation Implementation |
| RAM | : | Random Access Memory |
| CSS | : | Cascading Style Sheets |

# 1. INTRODUCTION

In the area of computer science, RPA can be described as the process of replacing human repetitive tasks with automated software implementations [1]. In other words, it is a "set of techniques concerning the operation and use of automata (robots) in the execution of multiple tasks in place of man" [2].

RPA is considered as relatively a recent technique tool that consists of software elements (bots) that imitate the manual tasks conducted by humans via different computer applications. The tasks to be automated are usually rule-based, well-structured, and repetitive. Robots can do a variety of functions, including screen scraping, automated mail query processing, and compiling payroll information from several sources. [3].

According to [3], 42 literatures outlined a number of advantages that can be attained with the use of RPA, including a decrease in the amount of time and energy needed to do manual operations, which saves on both. This has been demonstrated to be between 30% and 70% in terms of waiting times, job handling times, and process cycle times.

The exploration of the RPA concept has gone up very fast in recent years, especially in the digital business sectors. According to the most recent forecast from Garner Inc., despite the COVID-19 pandemic impact on economic sectors, global robotic process automation software revenue was projected to reach $1.890 B in year 2021, an increase of 19.49% from year 2020 [4].

While the Artificial Neural Network (ANN) can be defined as a model created to mimic how the brain carries out an interesting activity or function [5]. The primary goal of the ANN concept is to develop a network of interconnected components known as neurons to carry out complex activities and functions that cannot be carried out by computer programming or RPA.

Technology has advanced to the point where classification, grouping, pattern recognition, and prediction are now frequently used across a wide range of disciplines. Additionally, in machine learning (ML), ANNs are now competitive with traditional statistical and regression models in terms of usefulness [4].

Although RPA has many benefits and advantages, it also has limitations due to that the bots in RPA can only follow logical rule-based processes. As they do not have the ability to extract patterns from data or recognize images or speech. Hence, RPA programs its blot to process functions without understanding these functions pattern and logic behind them [6]. Effective modelling of the procedures necessary in instructional activities can be used to overcome these restrictions. Instead of only attempting to duplicate people's everyday operations, RPA-ANN uses the advantages of AI to learn how people make decisions and carry out complicated jobs. [7].

According to the research discussed in [3], the majority of real-world RPA implementations are used for organizational business objectives, whereas the research community hasn't looked into the use of the RPA concept in teaching activities. Additionally, the majority of earlier research focused on applying ANN algorithms and tools to enhance business activities, but RPA's potential to enhance ANN processes was not thoroughly explored. As the question of “how to use the RPA tools effectively for teaching purpose in the ANN-based research?” has not been discussed.

Therefore, this work aims to present and investigate an approach of exploiting RPA model in improving the implementation process of teaching activities in online form.

The ANN-based academic research to be used in this work is a research work that aims to predict students’ performance based on their study and non-study activities (lunch time, test preparation course, math score, reading score and writing score).
For this purpose, the datasets are gathered from Kaggle resource [8]. This involves using the Probabilistic Neural Network (PNN) prediction model, which consists of a group of artificial neural networks that were built using PARZEN’S approach [9] to forecast users’ performance. Eventually, the prediction result is evaluated by comparing predicted results performance with actual scores.

In the upcoming chapters we will talk about the usefulness of RPA and how It is a time saver and a cost-effective one too. Important features of robotic process automation include platform independence, scalability, and intelligence and an explanation about Macro recordings and how they are so flexible and accurate that you can do almost anything with them, the definition of Machine Learning, what kind of algorithms it uses and how important it is. We will use ANN as a type of the dynamics of gene expression as well and how it was abbreviated from how the human brain works. That will be explained in detailly in chapter (2). In the chapter that follows (3), we discuss and talk about Automation Process and how to implement the manual tasks (in the majority of teaching activities) that are proposed for RPA automation certain steps and evaluate the implementation of the proposed RPA model in ANN research through some aspects. And in chapter (4) of course we will implement it using (KNIME)) which is an analytics platform workflow system providing a platform for connecting tools graphically and test it at the end of it. At the (fifth) chapter the results will appear when we do efficiency and time consumption tests. The rest of the chapters are the challenges (that came along the way while writing this paper), conclusion and references.

# 2. GENERAL INFORMATION

## 2.1 ROBOTIC PROCESS AUTOMATION

Robotic process automation (RPA) is a technology that makes it simple for all users to automate digital operations. Robotic process automation significantly reduces the amount of time and labour required from humans. [reference] [10]. It not only saves time, but also money. Robotic process automation has the benefit of being independent in terms of platform, scalable, and intelligent [11]. RPA is used to automate different supplement chain processes, with data entry, predictive repairment and after-sales service support. RPA is used across organizations to automate hard rote tasks. RPA is used by telecommunications corporation to figure brand-new services and the related billing structure for modern accounts [12].

Robotic process automation (RPA) is a technology software that makes building, deploying, and managing software robots very easy, it emulates human beings' labour linked with digital systems and software. Just like humans, software robots can do matters like the understanding of what's presented on a certain screen, finish wanted keystrokes, system navigation, identify, extract data, and preformation of high range known activities. Unlike software robots that can do it more consistently than humans and much quicker, and there's no need to get up and stretch or take a break [13].

(Important features of robotic process automation include platform independence, scalability, and intelligence it also provides extra security, especially for sensitive data and financial services.)

RPA in business helps firms become more profitable, adaptable, and responsive by streamlining procedures. By reducing menial duties from their workdays, it also boosts employee satisfaction, engagement, and productivity [14].

RPA can be quickly adopted and is non-intrusive, which helps accelerate the digital transformation. Additionally, it's excellent for automating processes using outdated systems that don't have access to databases, virtual desktop infrastructures, or APIs (VDIs) [15].

RPA technology system is modifying how mankind is getting work done.

Contrary to people, software robots perform routine, low-value tasks such as entering into programs and systems, moving files and folders, copying, removing, and re-inserting data, filling out forms, and completing repetitive analysis and reports. Even cognitive tasks like translating text, participating in chats and conversations, comprehending unstructured data, and applying cutting-edge machine learning models to make difficult decisions can be performed by sophisticated robots [16].

Humans may concentrate on the activities they excel at and find most enjoyable when robots perform these types of routine, important, high-volume tasks, including as innovation, collaboration, creation, and customer interaction. Businesses gain a lot from this: they are more productive, efficient, and resilient. It makes sense that RPA is changing the history of the workplace.

Presently, across a wide range of businesses and activities, RPA software system is generating new efficiencies and freeing workers from monotonous repetitive work. Businesses have used RPA in a variety of industries, including finance, compliance, law, customer service, operations, and information technology. These companies are situated in a variety of sectors, including manufacturing, public sector, retail, healthcare, and financial services. Additionally, that is just the start. [13]

RPA is spread worldwide very extremely for obvious reasons which are that it is broadly applicable. Any large, iterative process that is governed by business rules is a fantastic candidate for automation, and cognitive processes that demand higher-order artificial intelligence capabilities are increasingly candidates as well.

Technology that can do far more than just assist with the automation of a particular process is required in order to create and manage an enterprise-wide RPA program. You need a platform that can assist you in building and managing a new enterprise-wide capability and enabling you to become a fully automated enterprise. You need end-to-end support from your RPA technology to find amazing automation possibilities everywhere, construct high-performing robots quickly, and manage hundreds of automated workflows. [10]

It's easy to understand why RPA has gained popularity all over the world when you combine its straightforward implementation with its quantitative advantages over other enterprise technology.

RPA is helpful in many various sorts of sectors since it helps them manage their own operational problems in new and simple ways.

Heads of functional areas from customer service to finance to human resources to marketing and beyond sees that RPA improves many processes, produces higher capacity, fewer errors for key processes and faster output.

From a CFO's standpoint, investing in RPA technology makes more sense because it offers a quick return on investment and necessitates less upfront expenditure than other business technology.

The simplicity with which software robots may access and function within legacy systems is one of the many reasons IT leaders think RPA can be adopted with minimum interruption. Modern RPA technology offers platforms that are scalable and appropriate for the enterprise because RPA has emerged as a key enabler of the digital transformation. Employees may use robotic assistants in the office with relative ease, and the low-code nature of RPA makes it possible for them to become citizen developers capable of developing their own simple automations. [17].

There are many toolkits to use the process of recording steps, including UiPath, Ui.Vision, but they are not free, unlike the macro record program, which provides free and easy to use platform.

## 2.2 MACRO RECORDER

By using the automation technique known as macro recording, you can record keystrokes and mouse movements from the screen and utilize them to create a script. With macro recordings' versatility and accuracy, you can achieve almost anything. Create a macro that pastes a template, opens the current email, and opens a new one. To automate reports with a keyboard shortcut, create macros that launch applications, alter formatting, save files, and export data into a worksheet. Even macro scripts that open webpages, open text fields, open drop-down menus, and submit forms can be written. identical window sizes and positions [18]

Macro Recorder additionally records the location and dimensions of any program windows that pop up while the recording is being done. Macro Recorder restores the window widths and locations during playback to guarantee that the macro may be appropriately replayed each time. Furthermore, the clever algorithms may transform erratic mouse movements into attractive curved or linear designs. Because there are no visual interruptions and the playback speed may be changed for specific mouse motions or all mouse movements, this is excellent for making screencasts [19]. movements. Only clicks are conducted when a mouse movement is removed from playback. Automation is made simple for everyone—not just programmers—by the Macro Recorder. Additionally, Macro Recorder eliminates the need to learn a complex scripting language because everything is done through the user-friendly interface. The macro recorder program is shown in Figure 2.1 below.

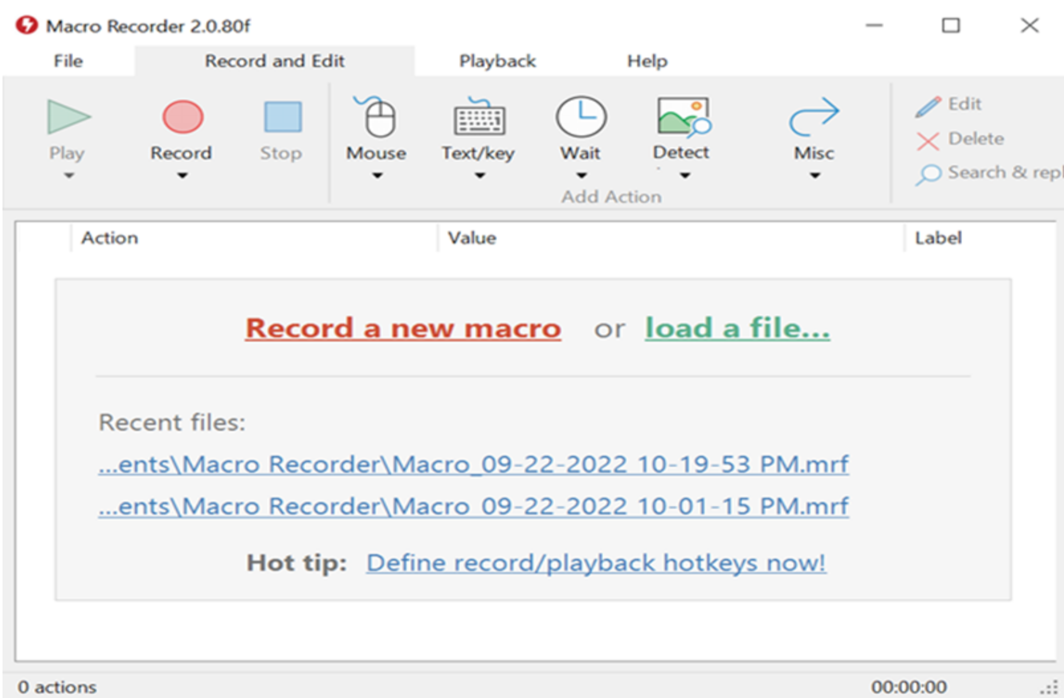



**Figure 2.1:** Macro record software.

You can use Microsoft Word’s Equation Editor or Math Type to write equations in Word. Equations that are inserted in the text as screenshots will be rejected. Below you may find an example of how to use an equation.

### 2.3 MACHINE LEARNING

With regards to artificial intelligence, IBM has a lengthy history. With his studies on the game of checkers, one of its own, Arthur Samuel, is credited with coining the term "machine learning" (PDF, 481 KB) (link lives outside IBM). A self-described checker’s master named Robert Nealey faced up against an IBM 7094 computer in 1962, but he lost. Although it may not appear noteworthy in light of what is now feasible, this accomplishment is usually regarded as a significant advance in the field of artificial intelligence. Machine learning is a branch of computer science and artificial intelligence (AI) that focuses on using data and algorithms to simulate how humans learn while continually improving its accuracy. And that enables software programs to forecast outcomes more accurately without being specifically trained to do so. In order to forecast new output values, machine learning algorithms use historical data as input. [20] Machine learning is particularly important since it helps companies develop new products and provides them with a picture of trends in customer behaviour and business operating patterns. Machine learning plays a crucial role in the daily operations of many of the world's most successful companies today, including Facebook, Google, and Uber. Machine learning has become a significant competitive advantage for many businesses. [21] Traditional machine learning is frequently categorized by how a prediction-making system learns to increase its accuracy. There are four basic learning methods: supervised learning, unsupervised learning, and semi-supervised learning. The kind of algorithm used depends on the type of data that data scientists want to predict [22].

Machine learning is employed in many different applications nowadays. The recommendation engine that drives Facebook's news feed is arguably one of the most well-known applications of machine learning. [23] When it comes to perks, machine learning can aid businesses in better comprehending their clients. Machine learning algorithms can discover associations and assist teams in customizing product development and marketing campaigns to customer demand by gathering customer data and comparing it with behaviors over time. Some businesses base their business models primarily on machine learning. For instance, Uber matches drivers with riders using algorithms. Google surfaces the ride adverts in searches using machine learning. But there are drawbacks to machine learning. It can be costly, first and foremost. Data scientists, who earn significant salaries, are often the ones in charge of machine learning projects. These initiatives also call for costly software infrastructure [24].

Machine learning relies on input, such as training data or knowledge graphs, to comprehend things, domains, and the connections between them, much to how the human brain acquires information and understanding. Entities must be defined before deep learning can start [25].

A model is developed using machine learning algorithms on a training dataset. The created model is used by the trained ML algorithm to make predictions as new input data is supplied [26].

A Decision Process: Typically, predictions or classifications are made using machine learning algorithms. Your algorithm will generate an estimate about a pattern in the input data based on some input data, which can be labelled or unlabelled. A function for errors: The model's prediction is evaluated using an error function. If there are known examples, an error function can compare them to gauge the model's correctness. [27] A model optimization process: Weights are altered to lessen the difference between the known example and the model estimate if the model can match the data points in the training set more accurately. This "evaluate and optimize" procedure will be repeated by the algorithm, with weights being updated automatically, until a predetermined level of accuracy is reached [27].

## 2.4 NEURAL NETWORK

As a human being, the recognition of the individual letters is called pattern recognition, the icons or symbols on a certain page being the patterns that the brain need to recognize. For example, while you are reading this paper your brain is going to sort out the signals that it is being sent by your eyes so that it can recognize the letters on this page and gather them all together into words for you to read or sentences or paragraphs and so on. Yet, for a printed book, the letters are all in a specific type, where all the letter 'a's to some extents are identical. Machines are designed to identify the letter ‘a’ easily, the reason for it is there is only one patter to be recognized. Now if we want to design a machine that is able to identify hand-written patterns what is the action to be done if we ask ourselves. a neural network consists of interconnected components. These elements were inspired from biological nervous systems studies. In some other words, these studies aim to build machines that can work like a human brain to achieve this we need to use components that act similar to biological neurons. Neural network’s main function is if given a valid input pattern in return you will receive an output pattern. This is only a summary of how neural network works, one of many neural network operations will be detailly described this operation is called -pattern classification which is identified as it is the procedure of sorting patterns into a group or another [1].

The artificial neural network will serve as a model for the kinetics of gene expression in our research. The significance of a single gene product's regulatory impact on the expression of other genes in the system is defined by a weight matrix. The approach sees positive and/or negative responses as part of multigenic control. A single network with two networks connected together that model transcription and translation independently describe the entire gene expression process. Many of the networks of interacting parts that make up natural processes influence one another's states throughout time. The linkages between them and the updating of the rules for each element are key to their dynamism. Genomic regulatory networks are just one of many examples of networks of this type. Each of these processes will be represented by a unique network that is guided by a unique weight matrix. There are procedures for computing parameters for experimental data that are modelled, and these procedures will be discussed [2].

The study discusses numerous applications of ANN techniques in various fields, such as nanotechnology, science, agriculture, climate, computing arts, medicine, engineering, environmental, mining, technology, and business, among others. It also offers a classification of ANNs and informs the reader about current and emerging trends in ANN applications research as well as the areas of focus for researchers. Also, it provides and evaluates challenges, contributions, concurrent performances, and the ANN application evaluation process. Even artificial neural network techniques like feedforward and feedback spread have demonstrated to function significantly better when applied to human difficulties. Based on data analysis parameters such processing, accuracy, speed, latency, fault tolerance, volume, scalability, performance, and convergence, we advised feedback and feedforward spread ANN models for study focus. Instead of adopting a single technique, we propose that future research concentrate

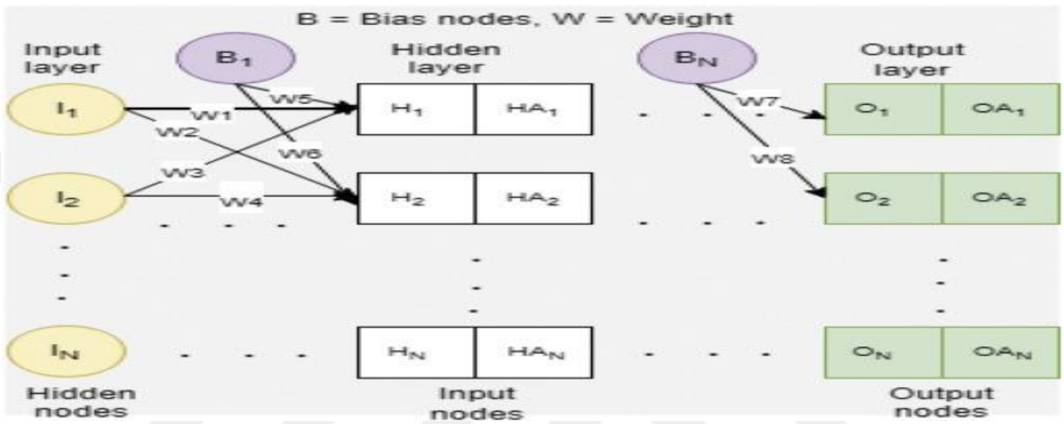



**Figure 2.2:** Structure of ANN.

on combining ANN models into a single network-wide application. [3] Since the layers of a neural network (NN) are independent of one another, any number of nodes may be present in a given layer. The term "bias node" refers to this random assortment of nodes. The bias nodes are always set to 1. [3].

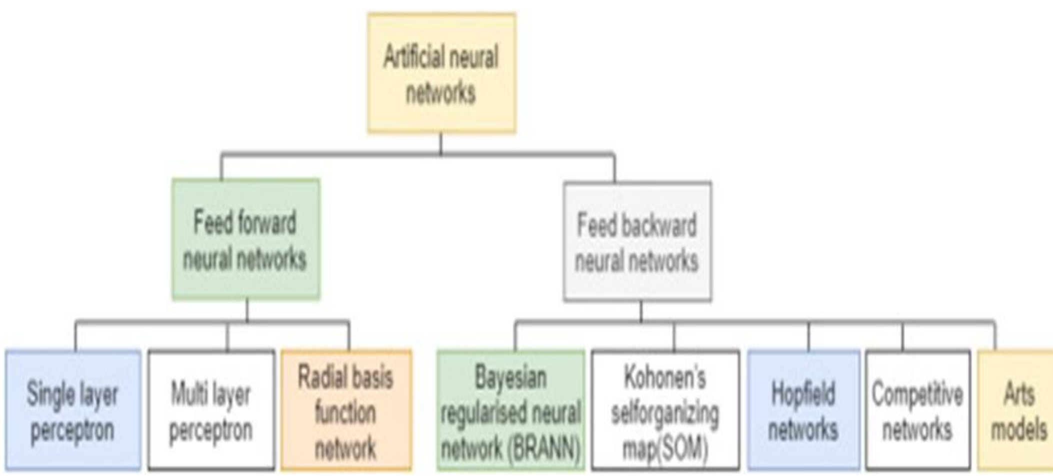



**Figure 2.3:** Types of ANN.

Using layers that are organized similarly to the processing units found in human neurons, feedforward neural networks are a type of machine learning classification technique. Each unit within a layer is linked to each and every other unit throughout the layers [3]. The three input units in Fig. 2.4, which are depicted as circles, are not a part of any layer in the network system and are frequently thought of as virtual layers with zero layers. Fig. 3 features one hidden layer and one output layer, which together display all the connections between the layers' components. A hidden layer is neither an input nor an output layer [3].

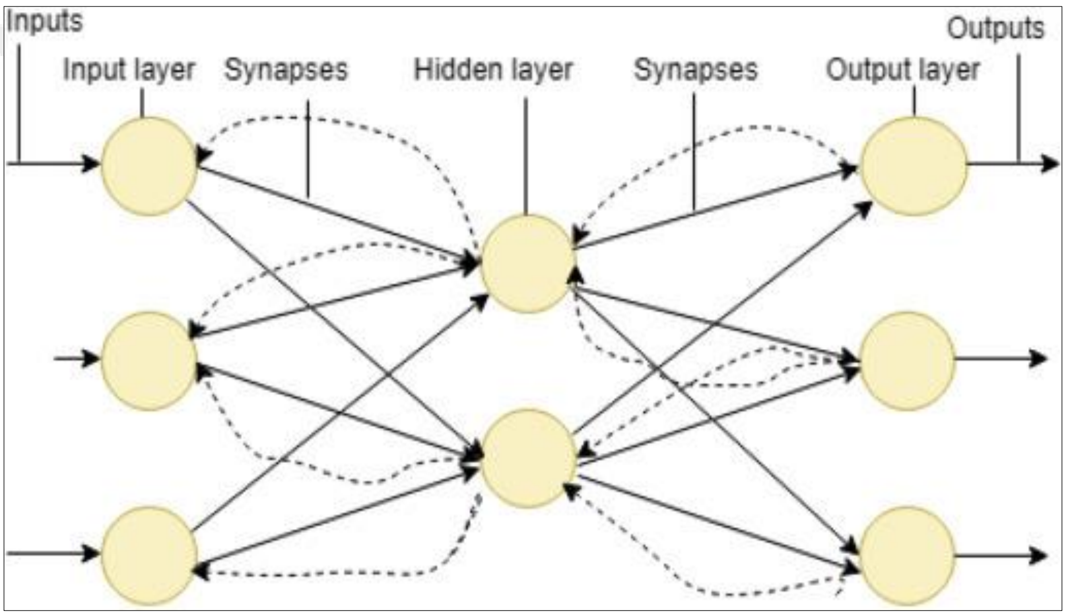



**Figure 2.4:** Layers in ANN.

A coordinated graph was formed in series by node connections that fed back on each other. Feedback NNs are able to exhibit dynamic terrestrial behaviour for a time period thanks to the coordinated graph in sequence [3].

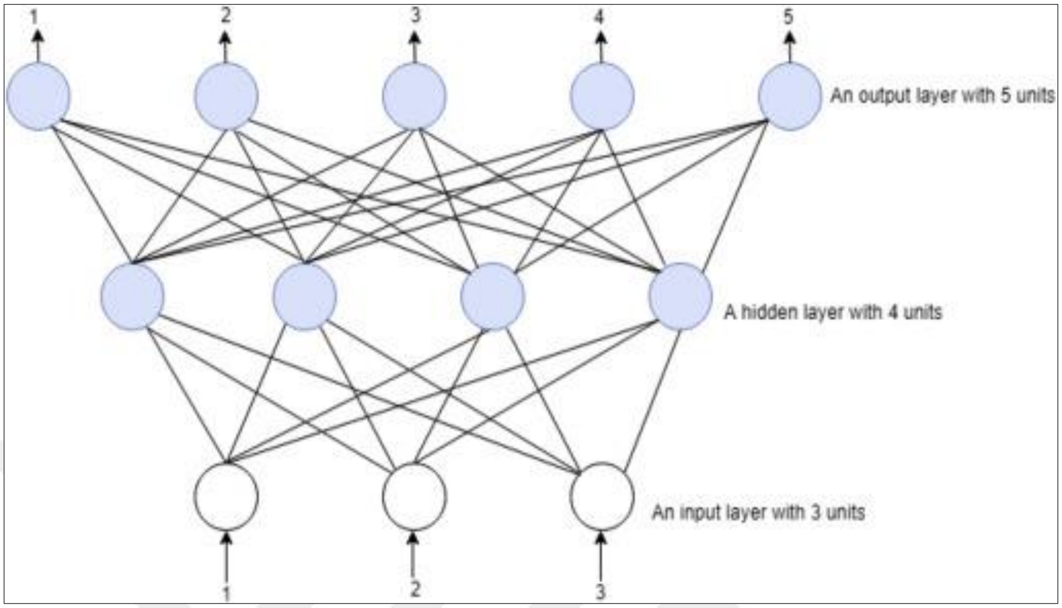



**Figure 2.5:** Layer connections in ANN.

# 3. METHODOLOGY

As presented before, the use of RPA concept has not been exploited effectively in the area of online teaching. Therefore, this work investigates the exploitation of RPA methodology in the area of lecturing activities. This is conducted by presenting an approach of applying the automation process of RPA in the lecturing and teaching process. This involves determining the manual processes in teaching area that can be automated using RPA.

## 3.1 AUTOMATION PROCESS

The manual tasks (in the majority of teaching activities) that are proposed for RPA automation can be listed in the following sequential steps:

i. Teaching activities cloud workload: in the teaching activities computer research, the datasets used in the investigation process are usually gathered from different cloud resources. And the most common steps in the gathering process involve the tasks of downloading and storing datasets from these resources into specific files in an organized manner. To save time and effort, we propose automating the data gathering process through exploiting RPA capabilities. In our prediction research, data gathering tasks determined for automation are:

   a. Download targeted datasets from the student performance datasets website.

   b. Saving these datasets in specific files.

ii. Data pre-processing (preparation): The second and the most important process in ANN-based research is the task of preparing and editing the gathered data sets to be used for the ANN model. This includes changing formats, filtering, data, extracting and many other different data mining applications. In most research cases, these tasks require a lot of repeated manual tasks that sometimes cause undesired mistakes and errors that affect research results. In our research case, the following tasks are assigned for automation.

a. We use the KNIME toolkit for neural networking which requires the data to be pre-processed and edited into (.csv) or (.xlsx) format. It's a platform for analytic report and integration that was developed by KNIME.com AG, in which different elements for data mining and machine learning are integrated and linked together in a form of workflow projects [28].

b. These data are also edited by omitting any information that is not used in the ANN prediction process to avoid overfitting in the neural network.

iii. ANN implementation: This part presents the implementation tasks of neural network models in the research work. Hence, all ANN research studies require using software toolkits for machine learning. Which includes the steps of starting and running the software tool used in the investigation process.

In our research case, the implementation steps are the following:

a. Starting the platform of KNIME toolkit.

b. Importing the targeted datasets into the KNIME platform.

c. Running the PNN prediction model to perform the investigation process.

It is important to note that the ANN software toolkit needs to be initialized manually to be ready for use in the RPA process.

iv. Results validation: this represents the activities and steps of evaluating the investigation results of ANN-based research. In most research work, this involves different tasks of using software toolkits assigned for results comparison, charting, exploring …etc. And because these activities are generally repeated steps, we propose to automate them with the use of RPA to improve the accuracy of evaluation.

In our research work, this was conducted by automating the validation of the predicting results through applying the scorer tools for the confusion matrix, which compares the clustering results with the actual user ID in the dataset.

Figure 1 shows the proposed approach of automating ANN-based research steps using the RPA automation method.

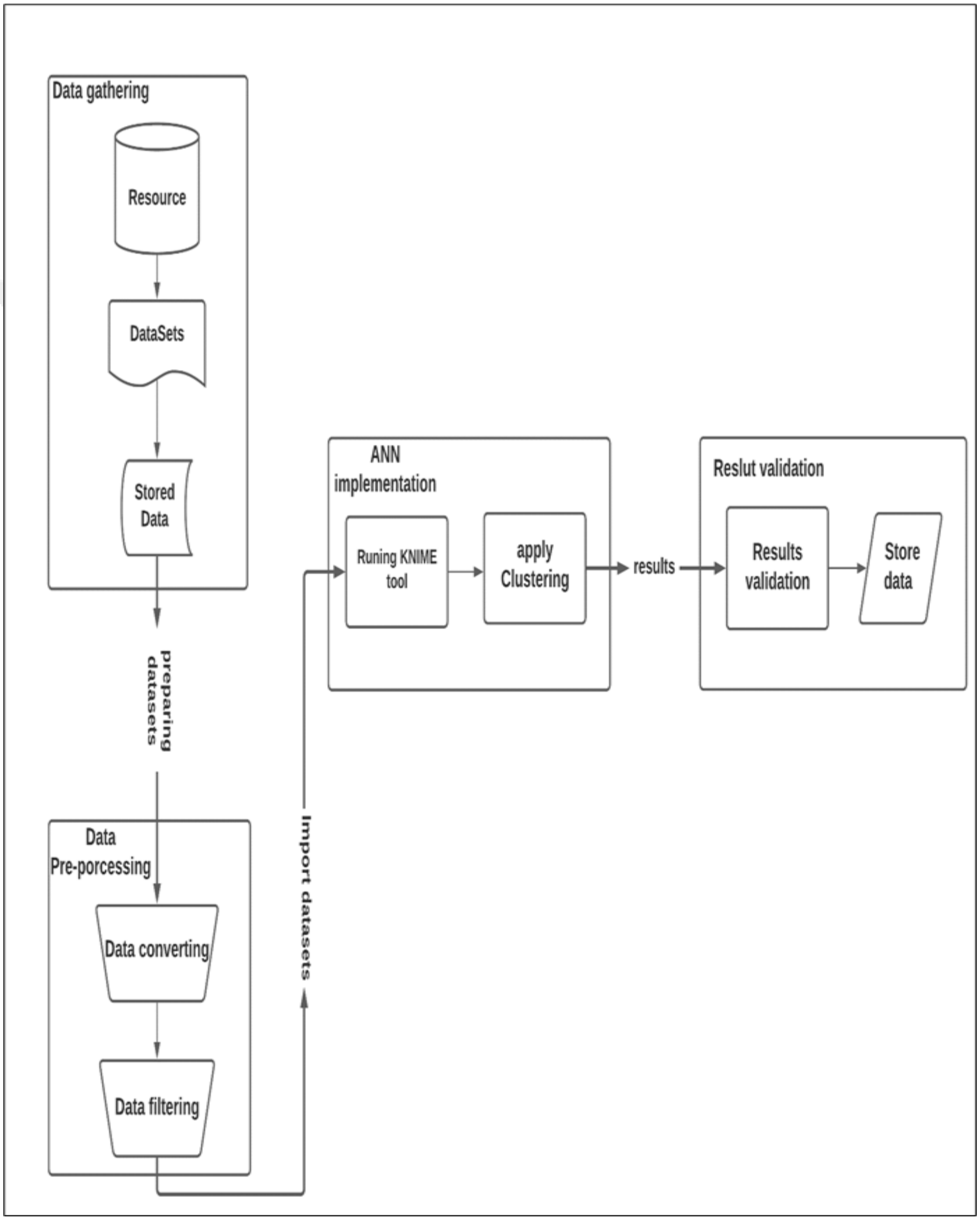



**Figure 3.1:** RPA model for ANN based research.

## 3.2 EVALUATION PROCESS

We evaluated the implementation of the proposed RPA model in ANN research through the following aspects of performance:

i. Time – Saving: This encompasses time consumption evaluation. This was conducted by comparing the time of implementing the research tasks using the RPA method with the same tasks conducted manually. This involves the time spent building the RPA model for sequences as well as the running time of the toolkit.

ii. Efficiency and accuracy: this include checking and measuring the accuracy and efficiency of implementing the investigation process of ANN-based research work when the RPA toolkit is used; and by checking the results of the implementation process.

# 4. IMPLEMENTATION

## 4.1 WORKFLOW DESIGN

The goal of the program is to predict the cloud user ID based on the requested time from workload traces such as Unilu Gaia. In order to conduct this, we design the KNIME workflow as the following parts (nodes).

### 4.1.1 Excel Reader

Despite just reading one sheet per file, this node can read one or more files simultaneously. String, integer, Boolean, date, and time are the only supported Excel kinds that may be read in; however, images, diagrams, etc. are not supported. Formulas may also be entered manually and revised as needed. Transformed data includes KNIME kinds including string, integer, long, double, Boolean, local date, local time, and local date time. Following execution, the node will scan the input file to determine the number and kinds of columns before creating a table with the automatically predicted structure and KNIME types. Performance of this node is limited (due to the underlying library of the Apache POI project). Reading large files takes a while and uses a lot of RAM (specifically xlsx files with enabled formula re-evaluation).

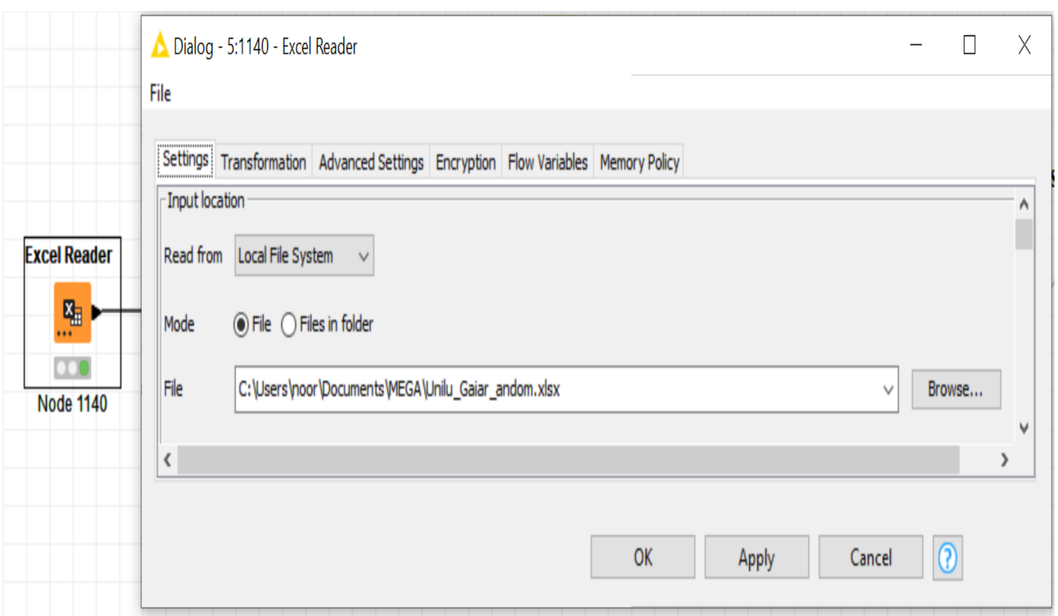



**Figure 4.1:** Excel reader node in KNIME.

### 4.1.2 Column Filter

Only the remaining columns from the input table are transferred to the output table after using this node to filter the input table's columns. Within the dialog, you can change the columns in the Include and Exclude lists.

Include: The names of the columns from the input table whose names should appear in the output table are contained in this list.

Exclude: Names of input table columns that should not be included in the output table are contained in this list. Filter You can filter the Include or Exclude list using one of these parameters to look for specific column names or name substrings.

Buttons: To shift columns between the Include and Exclude lists, use these buttons. All chosen columns will be moved by single-arrow buttons. All columns will move with double-arrow clicks (Consideration is given to filtering). If the input table specification changes, the existing exclusion list will be enforced to remain the same if this option is chosen. A warning is shown if any of the omitted columns are no longer accessible. (The list of included columns will automatically be updated as new columns are added.)

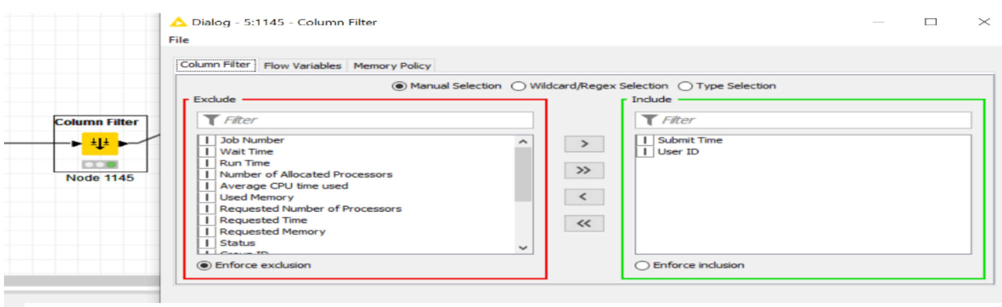



**Figure 4.2:** Column filter node in KNIME.

### 4.1.3 Normalization

All (numeric) columns' values are normalized by this node. You can select the columns you want to work on in the dialog. The dialog supports the following normalization techniques.

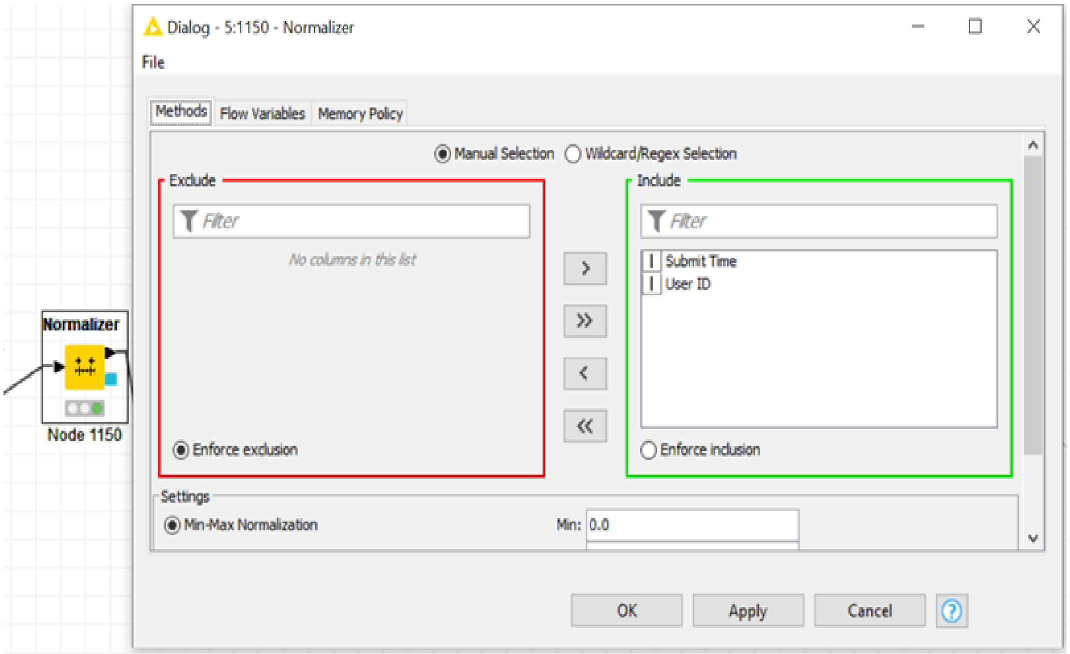



**Figure 4.3:** Normalization node in KNIME.

### 4.1.4 Partitioning

Two partitions are created in the input table (i.e., row-wise), for example, train and test data. The two output ports have access to the two partitions. The dialog box offers the following choices.

Absolute: Indicate how many rows there are in total in the first partition. If there are fewer rows than what is provided here, all of them are entered into the first table, and none are placed into the second table.

Relative: The proportion of the input table's total number of rows that are in the first partition. It must fall within a range of 0 and 100, inclusive.

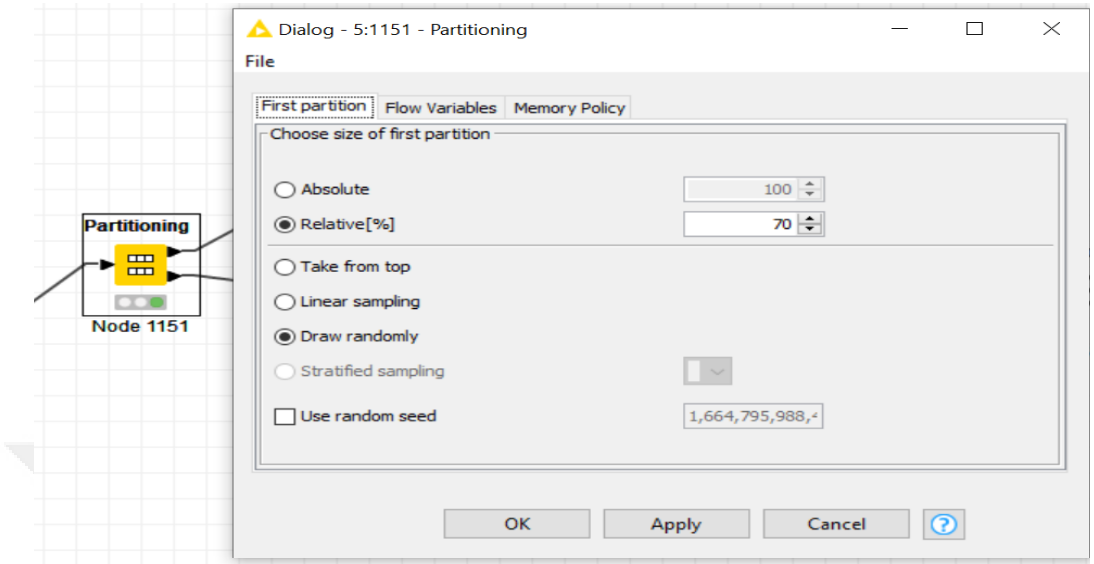



**Figure 4.4:** Partitioning node in KNIME**.**

### 4.1.5 PNN Learner (DDA)

Based on numerical data, this algorithm develops rules. Each rule is described as a high-dimensional Gaussian function that has been modified by the theta minus and theta plus thresholds in order to avoid conflicts with rules from other classes. Each Gaussian function is specified by a center vector (from the first covered instance) and a standard deviation that is modified throughout training to only include instances that are not in conflict. The input data's chosen numeric columns are used as training data, and extra columns are utilized as classification targets; either one column contains class information, or a selection of numeric columns with class degrees ranging from 0 to 1 can be made. The rules after execution and a number of rule metrics are included in the data output. The PNN model is contained in the model output port and can be used by the PNN Predictor node to make predictions.

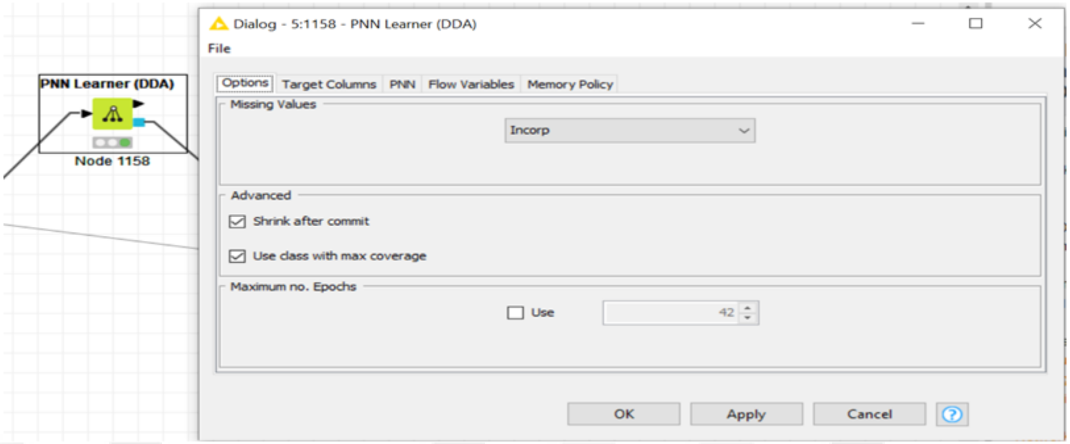



**Figure 4.5:** PNN learner node in KNIME.

Maximum number of epochs: If chosen, this option specifies how many epochs the algorithm will need to process the entire data set. If not, the algorithm will keep repeating this process until the rule model is stable, which is when no new rules have been committed and/or no shrink operations have been carried out.

Target Columns: Identify the target(s) to be classified. If more than one column (only numeric) is used, each column's values for class membership for each class specified by the column name must fall between 0 and 1.

Number of strings a column's (or a group of columns') worth of numbers is converted to strings. Remember to utilize the "Round Double" node if you need a sophisticated setting, such as rounding or representation in scientific notification.

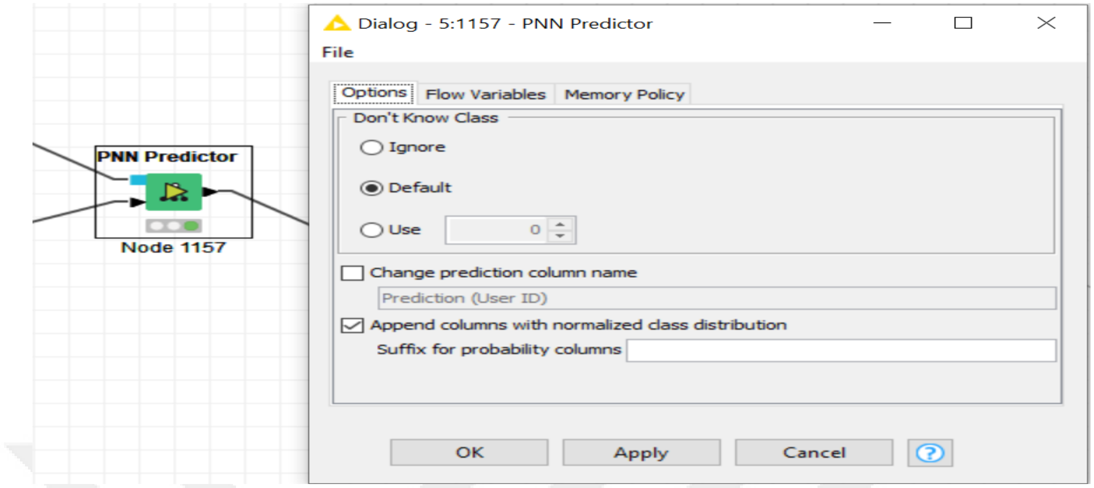



**Figure 4.6:** PNN predictor node in KNIME.

### 4.1.6 Scorer (JavaScript)

Demonstrates the confusion matrix and compares two columns by the pairwise values of their attributes. The values from the first selected column are represented in the confusion matrix's rows, while the values from the second column are represented in the confusion matrix's columns. There are three tables in the node's view, the first of which is the confusion matrix with the number of matches in each cell. The number of accurate predictions divided by the total number of records in the row or column yields row and column rates, which can be shown using a configuration setting. In order to pick the underlying rows, it is also feasible to highlight specific cells in this matrix. The selection can be passed to other JavaScript views. The second table reports a number of statistics specific to a given class such as True-Positives, False-Positives, True-Negatives, False-Negatives, Accuracy, Balanced Accuracy, Error Rate, False Negative Rate, Recall, Precision, Sensitivity, Specificity, F-measure. The last table contains overall statistics like the Overall Accuracy, Overall, Error, Cohen's kappa, the Correctly Classified and the Incorrectly Classified values. These three tables are also available as output ports of this node.

The node allows for personalized CSS style. In the node configuration dialog, you can easily condense CSS rules into a single text and set it as the flow variable "custom CSS." The list of available classes and a description of each may be found on our Documentation page.

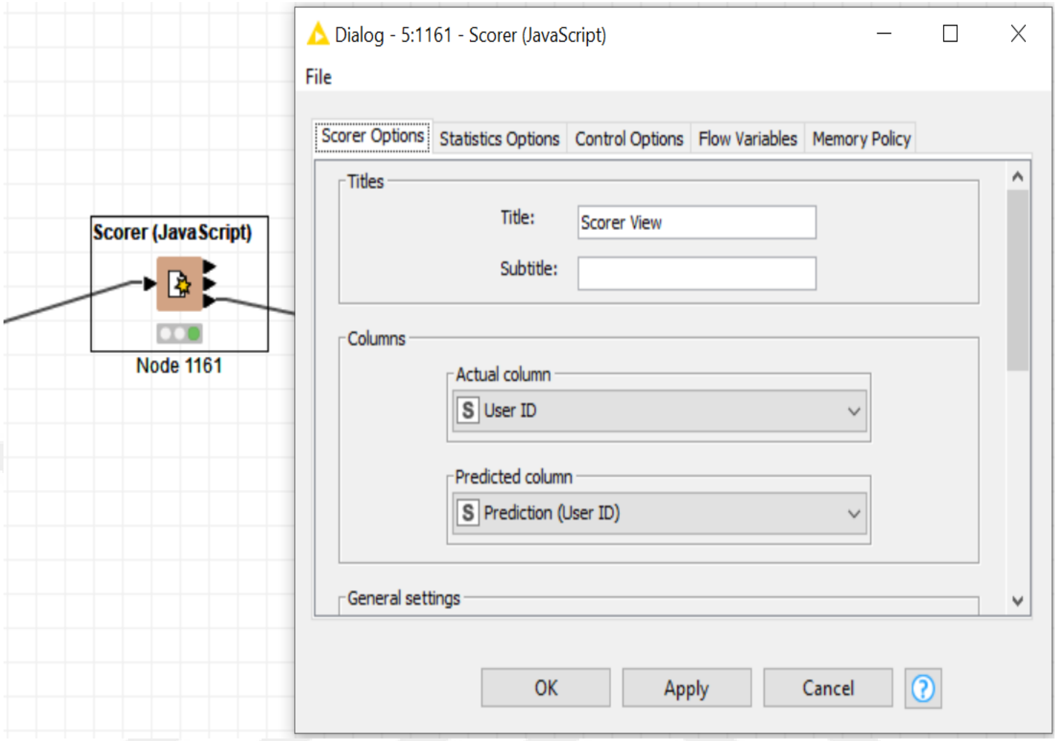



**Figure 4.7:** Scorer node in KNIME.

### 4.1.7 Excel Writer

The input data table is written by this node into an Excel file spreadsheet that can be read by other programs like Microsoft Excel. The node has the ability to add data to already existing Excel files or create whole new ones. The input data can be added as a new spreadsheet or after the last row of an already existing spreadsheet when appending. Multiple data table input ports allow the data to be written to or added to other spreadsheets inside the same file.

Two file extensions are supported by the node, and they are as follows:

i. XLS format: Prior to Excel 2003, this was the default file format that was utilized. In a spreadsheet of this type, the maximum number of columns and rows is 256 and 65536, respectively.

ii. XLSX format: Starting with Excel 2007, the default file format is the Office Open XML format. A spreadsheet in this format may have a maximum of 16384 columns and 1048576 rows, respectively.

Although appending is permitted, this node does not support writing files in the (.XLSM) format.

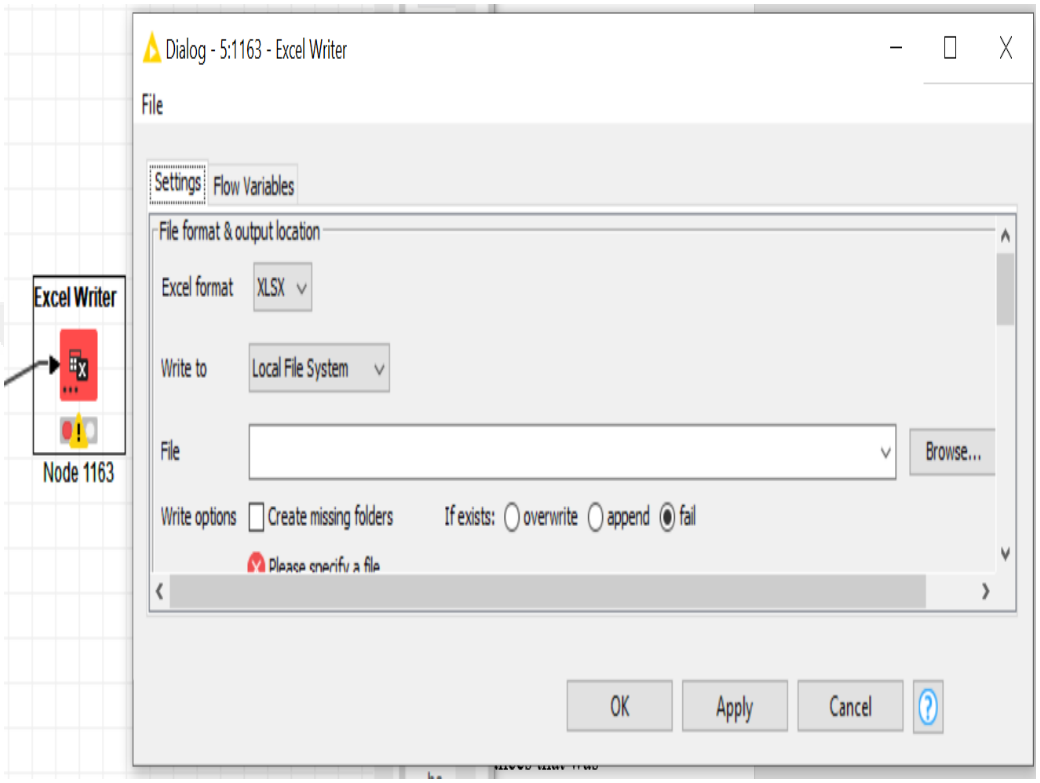



**Figure 4.8:** Excel Writer node in KNIME.

### 4.1.8 Testing

As presented previously, the dataset of cloud workload traces that was used in this work for the ANN prediction was gathered from the parallel workload archive website. In this test 500k.

Recorded from 5 traces were used. These traces are (Unilu-Gaia, ANL-interpad, KIT-FH2, METACENTRUM 2009 and RICC). Figure 4.10 below shows a sample of these traces format. Figure 4.10 shows a sample of this dataset.

| | A | B | C | D | E | F | G | H | I | J | K | L | M |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Job Number | Submit Time | Wait Time | Run Time | Number of Allocated Processors | Average CPU time used | Used Memory | Requested Number of Processors | Requested Time | Requested Memory | Status | User ID | Group ID |
| 2 | 7001 | 6298047 | -1 | 110 | 64 | -1 | -1 | 64 | -1 | -1 | 0 | 18 | -1 |
| 3 | 15028865 | -1 | 449 | 512 | -1 | -1 | 512 | 1800 | -1 | 0 | 96 | -1 | 3221 |
| 4 | 9331 | 7534176 | -1 | 0 | -1 | -1 | -1 | -1 | -1 | -1 | 5 | 20 | -1 |
| 5 | 16139706 | -1 | 10 | 8 | -1 | -1 | 8 | 1800 | -1 | 5 | 94 | -1 | 2927 |
| 6 | 7773 | 6568782 | -1 | 1 | 48 | -1 | -1 | 48 | -1 | -1 | 1 | 24 | -1 |
| 7 | 18393054 | -1 | 1823 | 128 | -1 | -1 | 128 | 1800 | -1 | 0 | 20 | -1 | 1636 |
| 8 | 18727311 | -1 | 0 | -1 | -1 | -1 | -1 | 864000 | -1 | 5 | 116 | -1 | 4203 |
| 9 | 15743749 | -1 | 9 | 8 | -1 | -1 | 8 | 7200 | -1 | 5 | 82 | -1 | 2849 |
| 10 | 9188467 | -1 | 6057 | 64 | -1 | -1 | 64 | 7200 | -1 | 1 | 52 | -1 | 1654 |
| 11 | 9420 | 7590203 | -1 | 13 | 16 | -1 | -1 | 16 | 3600 | -1 | 0 | 4 | -1 |
| 12 | 8704 | 7156623 | -1 | 38 | 512 | -1 | -1 | 512 | -1 | -1 | 5 | 36 | -1 |
| 13 | 2781 | 4744637 | -1 | 640 | 64 | -1 | -1 | 64 | 3600 | -1 | 1 | 4 | -1 |
| 14 | 15117263 | -1 | 11 | 128 | -1 | -1 | 128 | 1800 | -1 | 0 | 70 | -1 | 3253 |
| 15 | 8112175 | -1 | 1 | 8 | -1 | -1 | 8 | -1 | -1 | 1 | 2 | -1 | 186 |
| 16 | 7318 | 6448398 | -1 | 7 | 8 | -1 | -1 | 8 | -1 | -1 | 1 | 19 | -1 |
| 17 | 18817074 | -1 | 75 | 128 | -1 | -1 | 128 | 7200 | -1 | 1 | 110 | -1 | 4229 |
| 18 | 19738080 | -1 | 1765 | 1024 | -1 | -1 | 1024 | 7200 | -1 | 1 | 103 | -1 | 4163 |
| 19 | 16851263 | -1 | 536 | 1024 | -1 | -1 | 1024 | 7200 | -1 | 1 | 103 | -1 | 3599 |
| 20 | 2396 | 4270784 | -1 | 385 | 288 | -1 | -1 | 288 | 10800 | -1 | 1 | 3 | -1 |
| 21 | 9065 | 7474323 | -1 | 200 | 64 | -1 | -1 | 64 | -1 | -1 | 1 | 37 | -1 |

**Figure 4.9:** Sample of workload traces.

The gathering process includes downloading the dataset and convert it into (.csv) or (.xlsx) format. Then this dataset is imported in to the KNIME toolkit. In this toolkit, the prediction model is implemented via KNIME workflow nodes as shown in the Figure 4.11.

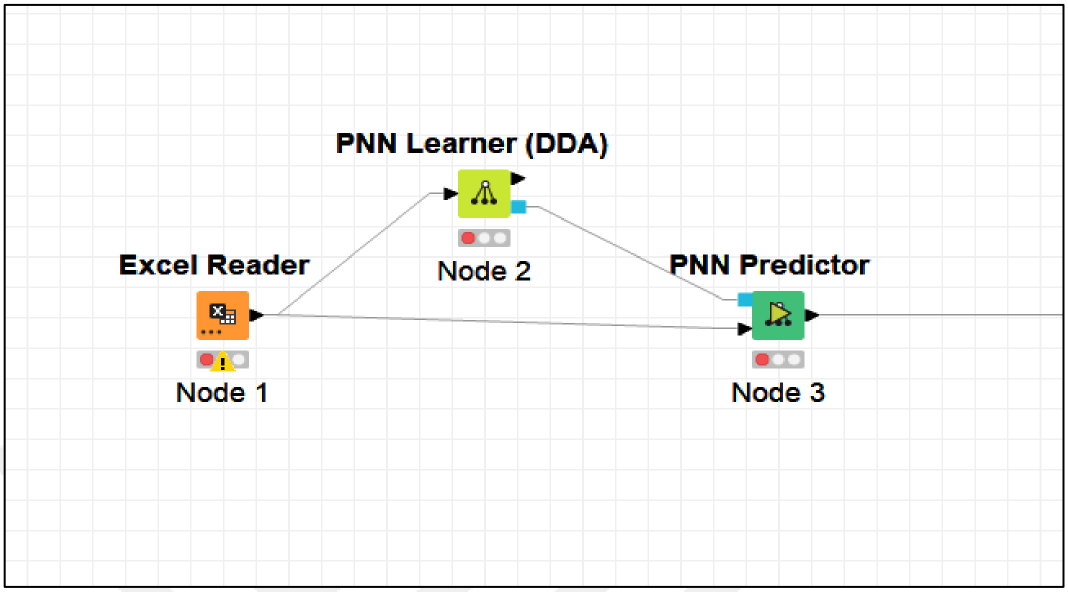



**Figure 4.10:** KNIME prediction model.

Afterword the implementation of RPA concept is conducted using Macro record toolkit, which is one of the most known tools for RPA implementations as presented previously. This platform provides the ability to automate computing tasks and steps through a sequence of programming blocks and workflows implementation in the form of elements. Each of these blocks and elements represents a task step that will be sequentially performed according to the user's design.

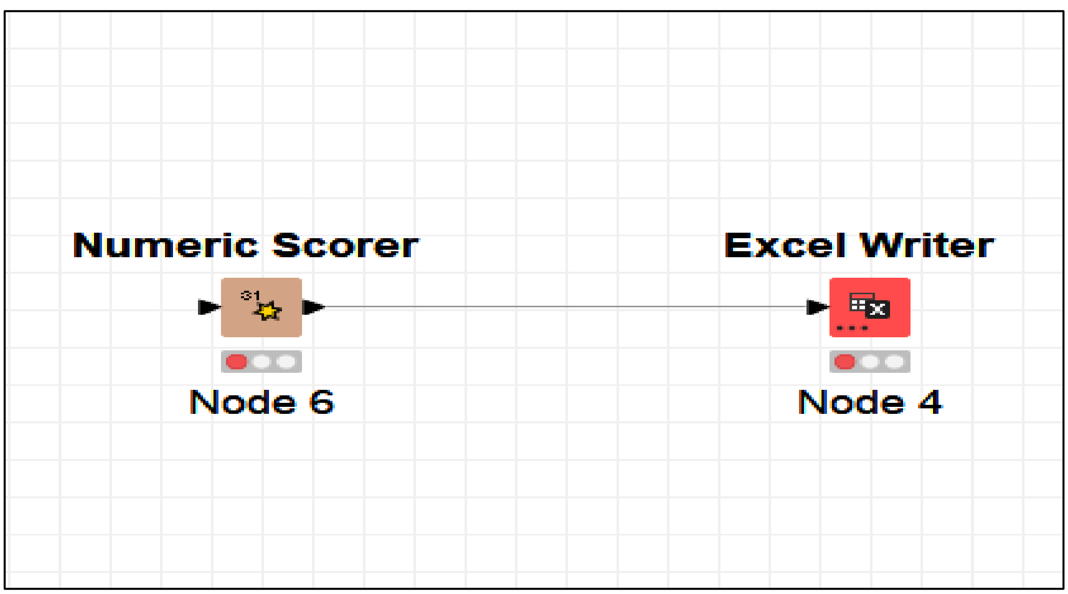



**Figure 4.11:** Result validation in KNIME.

Accordingly, we propose to automate the manual tasks of the RPA model shown in Figure 2.1 by using the Macro-record platform. This can be conducted by automating the sequential parts of (Data gathering, Data pre-processing, ANN implementation, and Results validation) in separate blocks using UiPath tools.

Firstly, the data gathering manual tasks are automated as shown in the sample in Figure 4.10. In this sample we automated the steps of downloading targeted datasets of User's log in cloud workload from Parallel Workload traces and saving these datasets into specific storage locations. This is followed by implementing the tasks of *Data pre-processing*, *ANN implementation*, and eventually *Result validation*.

It's important to imply that in our research testing the automation of all these tasks is conducted in a single connected automation process. However, for some research cases these tasks may require implementing the automation separately. This will be conducted by implementing and running the automation of each block shown in Figure 2.1 individually Then the result of each block is delivered to the followed one sequentially. The next task in the implementation process is the validation process, which consists of result evaluation and storing. These tasks are conducted in KNIME as shown in Figure 4.13.
To use the process of automation to solve problems in research by integrating the RPA with the neural network for the purpose of accurately measuring its ability to perform tasks

The macro record program used it to download data automatically, and the Knime program in this research. The Knime program runs a test where the data goes to the column filter part and every time it adds a new attribute (column) and then the process continues to the project part to expect the result as testing is done for it and then gives us the result. The result is stored separately for each experiment.

We run the Knime program and change the fields used in the production. Then we run the macro record program.

i. Run the program and open the site.

ii. The program goes to the user's website and downloads the data from it.

iii. It automatically downloads the data to register the website side, and then identifies the required data and loads it into the calculator. In order to use the program well we take care of the installation orientations and screen dimensions.

iv. We define the record and the program repeats the step each time by:

a. Data collection.

b. Makes the program apply the configuration change of the neural network by adding a new attribute. With every record we can specify the number of iterations where we test 30 times and every time it does the same additions. After downloading the data, the program performs data processing to prepare the data. Where the record can record the process easily (add all the steps recursively) and then store the result.and efficiency).

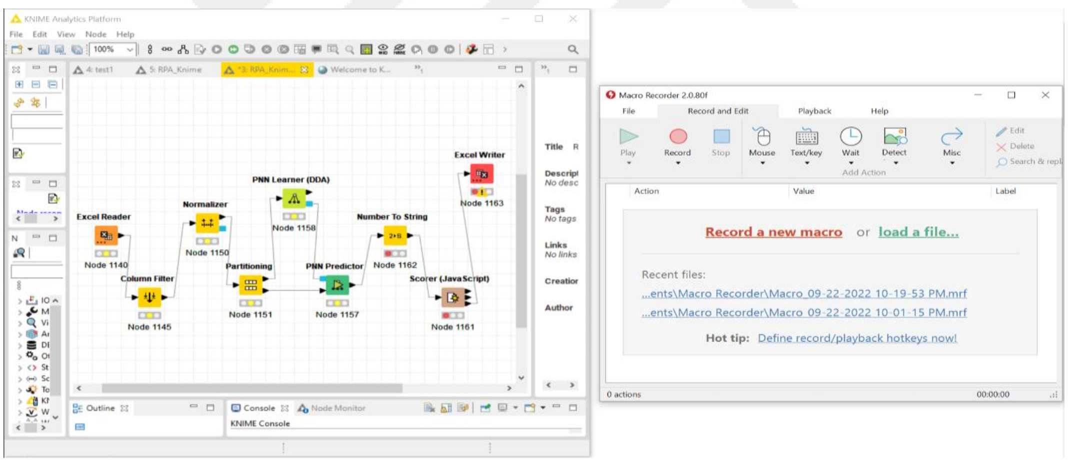



**Figure 4.12:** Sample of RPA-ANN automation using Macro-record (Data gathering).

# 5. RESULTS

The results of the RPN-ANN model performance evaluation are conducted in 5 tests. These tests represent the comparisons between the research steps and activities that were performed manually and the proposed RPA model with respect to the (Time consumption and efficiency).

i. Test 1:

The results of the first test show that the time consumption for manual *Data gathering* was 15 minutes (900 seconds) with 82% of accuracy and 960 seconds for Data Pre-processing with 84% of accuracy. While the manual implementation of both *ANN prediction model (PNN)* and manual *Result validation* (*Scorer node*) consumed less 7 minutes (420 seconds), with less than 85% accuracy. Accordingly, the overall consumption time for manual implementation was 38 minutes (2280 seconds).

On the other hand, by applying the RPA model, the time consumption of these activities was reduced dramatically to only 30 seconds to perform all steps with an average of 96% accuracy. These results are shown in Figures 5.1 and 5.2.

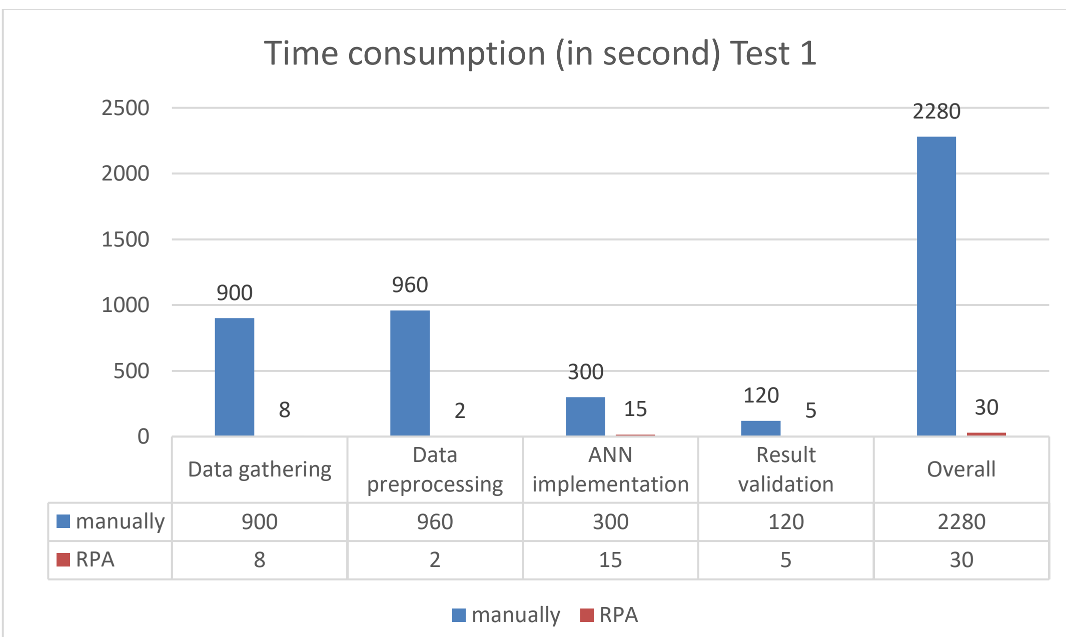
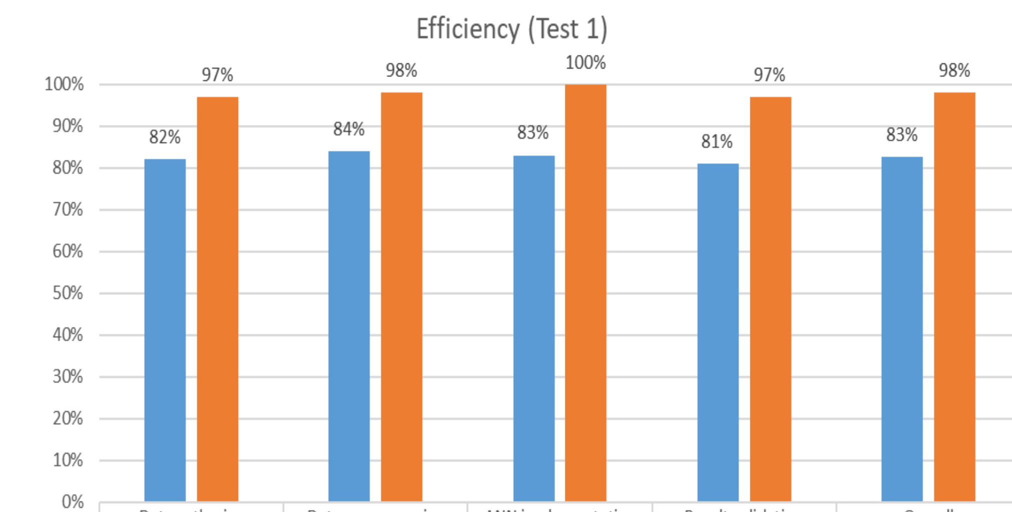



**Figure 5.1:** Comparison of Time consumption (1st test).




| | Data gathering | Data preprocessing | ANN implementation | Result validation | Overall |
|---|---|---|---|---|---|
| manually | 82% | 84% | 83% | 81% | 83% |
| RPA | 97% | 98% | 100% | 97% | 98% |

**Figure 5.2:** Comparison of Efficiency (1st test).

ii. Test 2:

In the second test, the results show that the time consumption for manual *Data gathering* was 33 minutes (1980 seconds) with 87% accuracy. This is followed by the *Data pre-processing* which consumed 900 seconds. While, the manual implementation of both *ANN* and *Result validation* consumed 6 minutes (360 seconds). These tasks were implemented with the accuracy of more than 85%. As shown in Figure 5.3, the overall consumption time for manual implementation was 33 minutes (1980 seconds).

By applying the RPA model (through Macro-recorder) with the same implementations, the overall time consumption of the same activities was reduced dramatically to only 32 seconds to perform all steps automatically with an average of 96% accuracy. These results are shown in Figure 5.4.

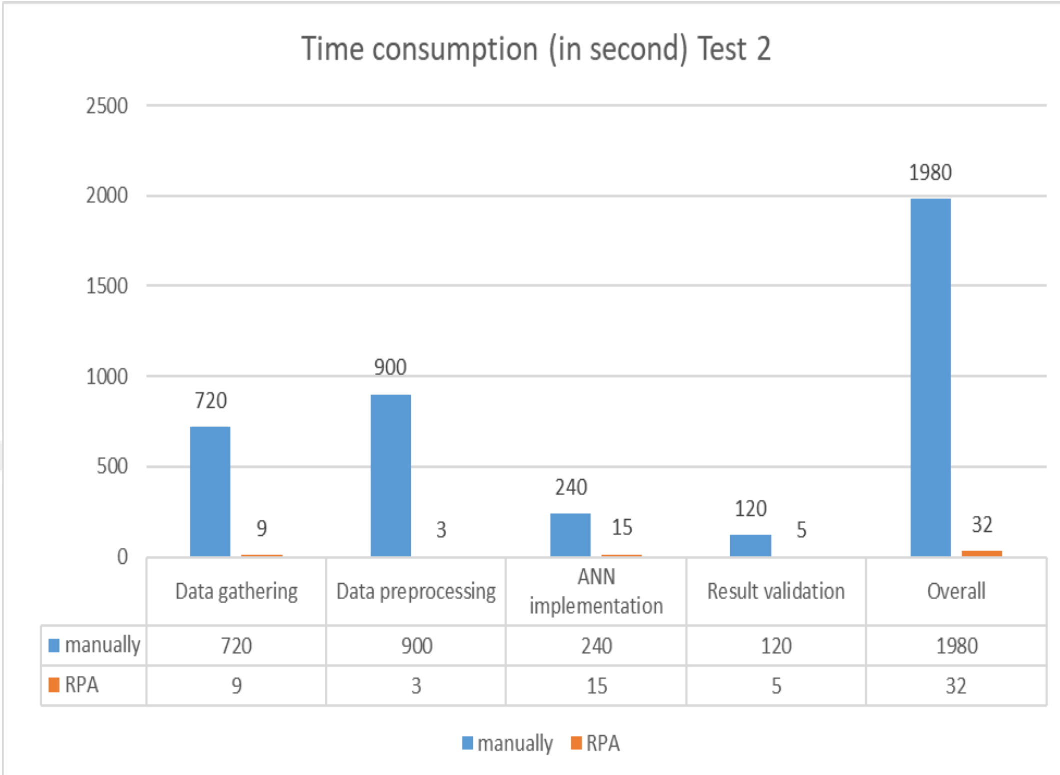



**Figure 5.3:** Comparison of Time consumption (2nd test).

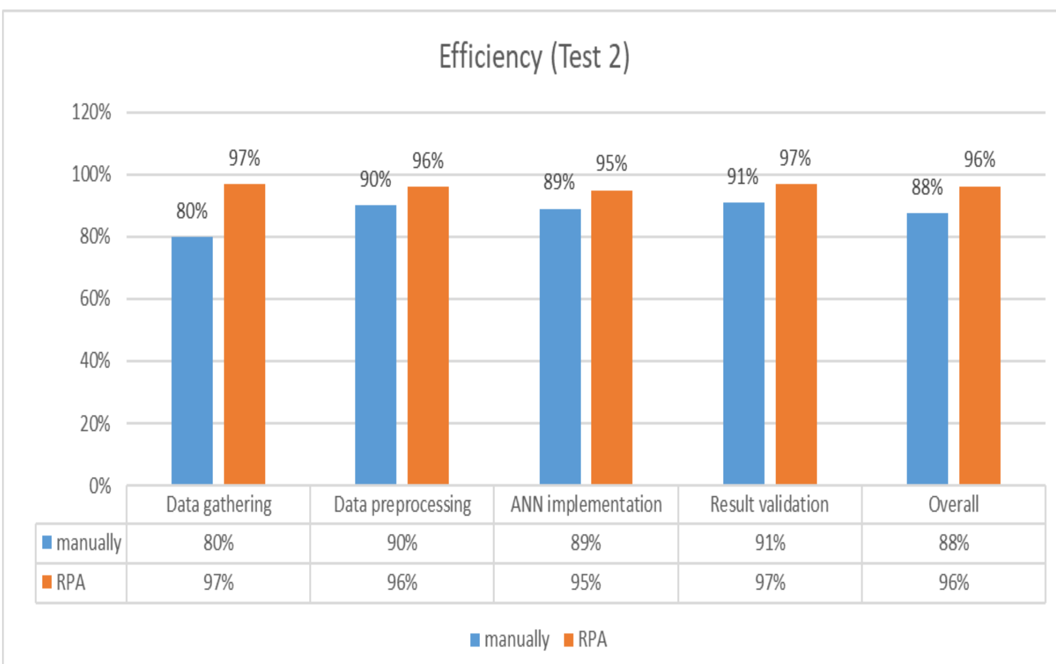



**Figure 5.4:** Comparison of Efficiency (2nd test).

iii. Test 3:

In the third test, the results of the comparison test show about 10 minutes (620 seconds) of the time consumption for manual Data gathering and (700 seconds) for Data pre-processing with the efficiency of 83% and 90% respectively. While the manual implementation of ANN and the Result validation of the prediction result recorded less than 6 minutes (360 seconds). These tasks were implemented with about 90% of efficiency (Accuracy). Accordingly, the overall consumption time for manual implementation was (1680 seconds) 28 minutes.

On the other hand, by applying the RPA model, the time consumption of the same tasks was reduced to 32 seconds to implement all these tasks with an average of more than 95% of efficiency. These results are shown in Figure 5.5 and 5.6.

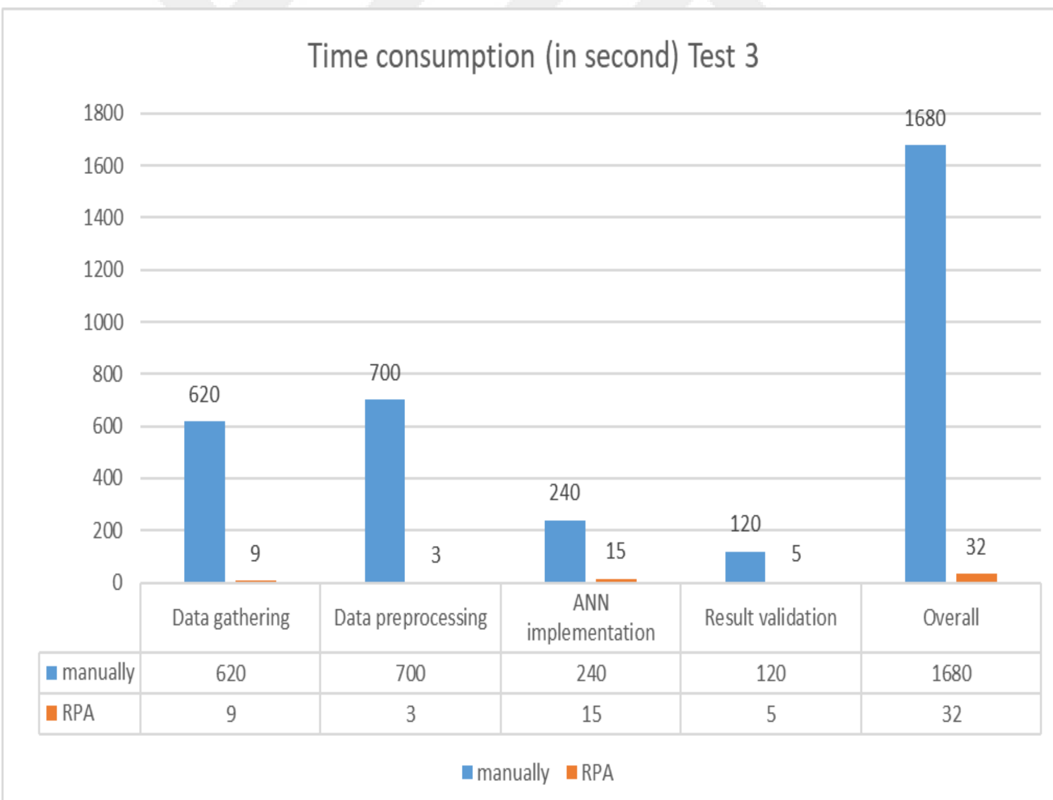



**Figure 5.5:** Comparison of Time consumption (3rd test).

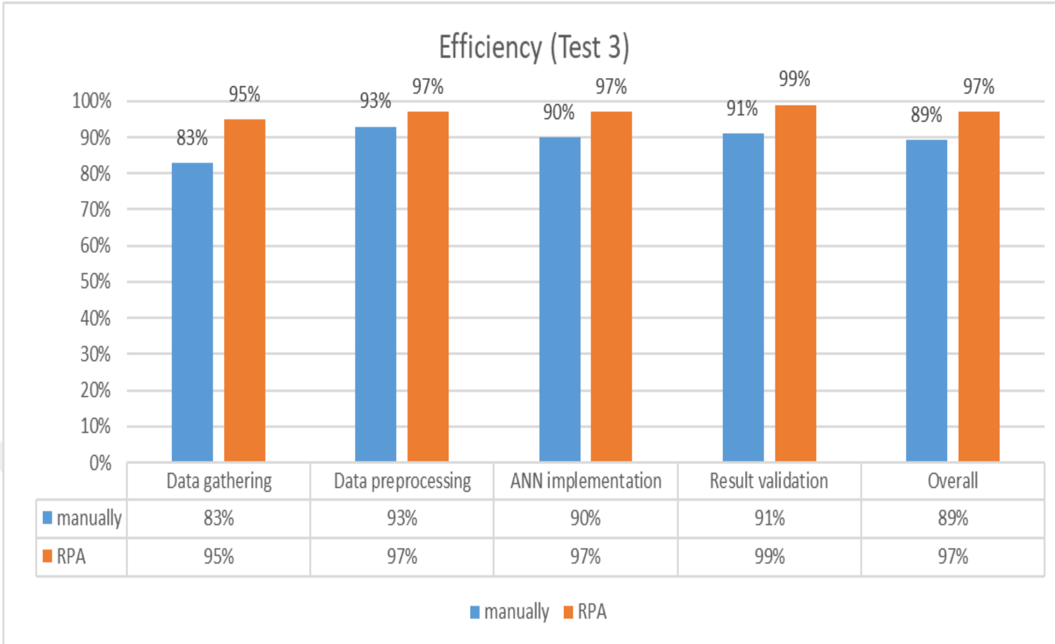



**Figure 5.6:** Comparison of Efficiency (3rd test).

iv. Test 4:

In the fourth test, the results of time consumption for manual tasks were as the following. For the *Data gathering* the time recorded for implementation was about 7 minutes with the accuracy of 87%. While both *ANN implementation* and manual *Result validation* recorded about 5 minutes (315 seconds), with the efficiency around 90%. The overall consumption time for the tasks in the manual implementation was around 32 minutes (1935 seconds) and the average efficiency of all the manual task was 90%.

In the comparison, by applying the Macro-recorder automation through the RPA model, the time consumption of the same tasks conducted above recorded 40 seconds to perform all steps with the average efficiency of 96%. These results are shown in Figure 5.7 and 5.8.

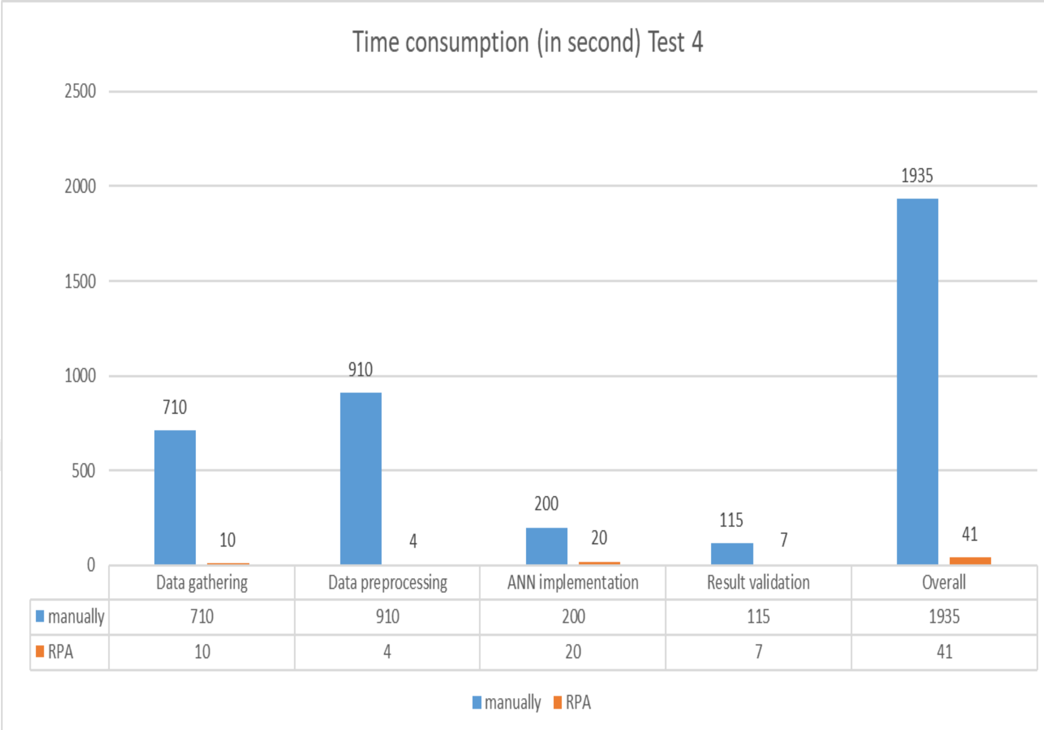



**Figure 5.7:**Comparison of Time consumption (4th test).

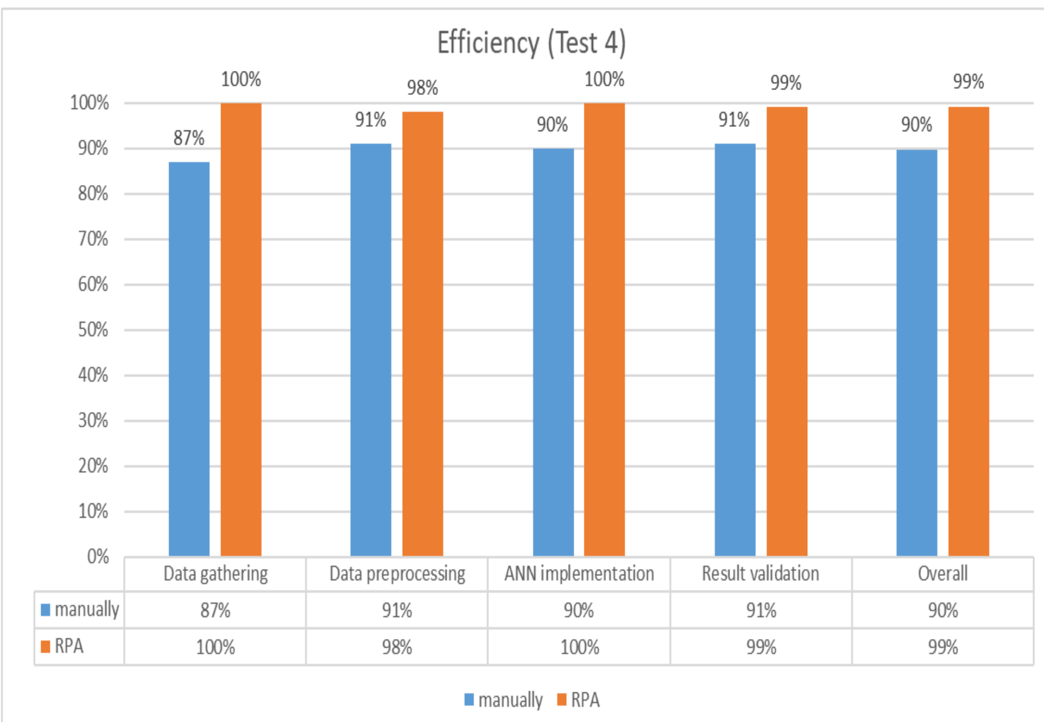
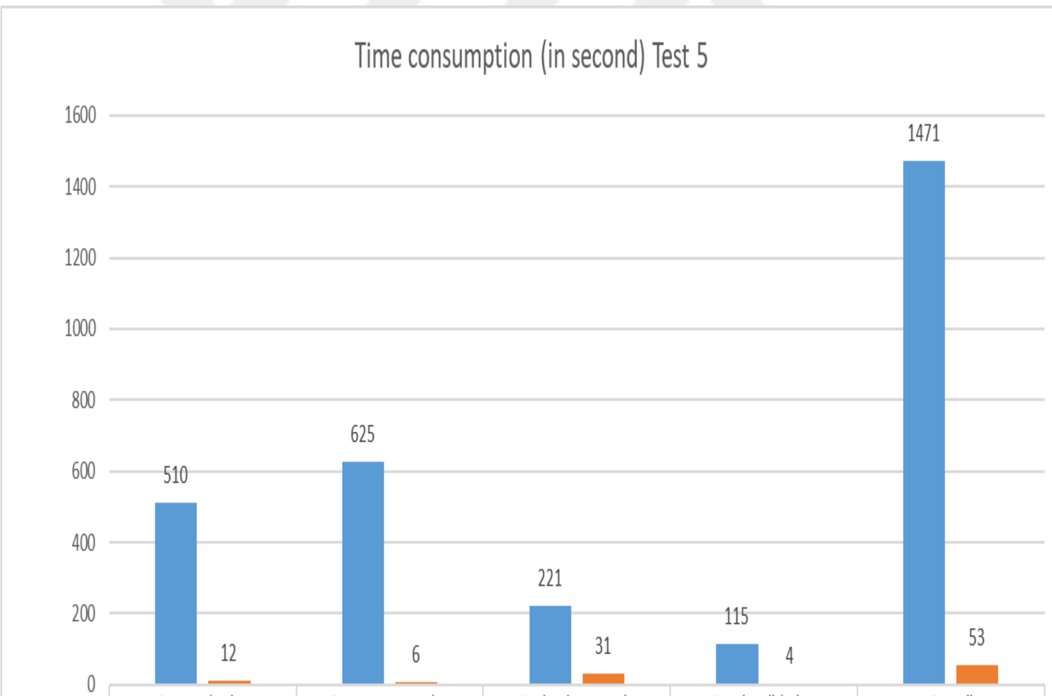



**Figure 5.8:** Comparison of Efficiency (4th test).

v. Test 5:

Finally, in the last test, the results show that the time consumption for manual *Data gathering* was about 8 minutes (510 second) with the efficiency of 73%. In addition, the Data pre-processing tasks recorded (625 seconds). While the manual performing of both *ANN implementation* and *Result validation* consumed around 5 minutes, with less than 90% of efficiency. Accordingly, the overall consumption time for manual implementation was 1470 seconds. And the efficiency (accuracy) of implementing these tasks manually was 81%.

On the other hand, by applying the RPA model of Macro-recorder, the time consumption of these activities was reduced dramatically to only 53 seconds to perform all steps with the efficiency (accuracy) of 98%.




| | Data gathering | Data preprocessing | ANN implementation | Result validation | Overall |
|---|---|---|---|---|---|
| manually | 510 | 625 | 221 | 115 | 1471 |
| RPA | 12 | 6 | 31 | 4 | 53 |

**Figure 5.9:** Comparison of Time consumption (5th test).

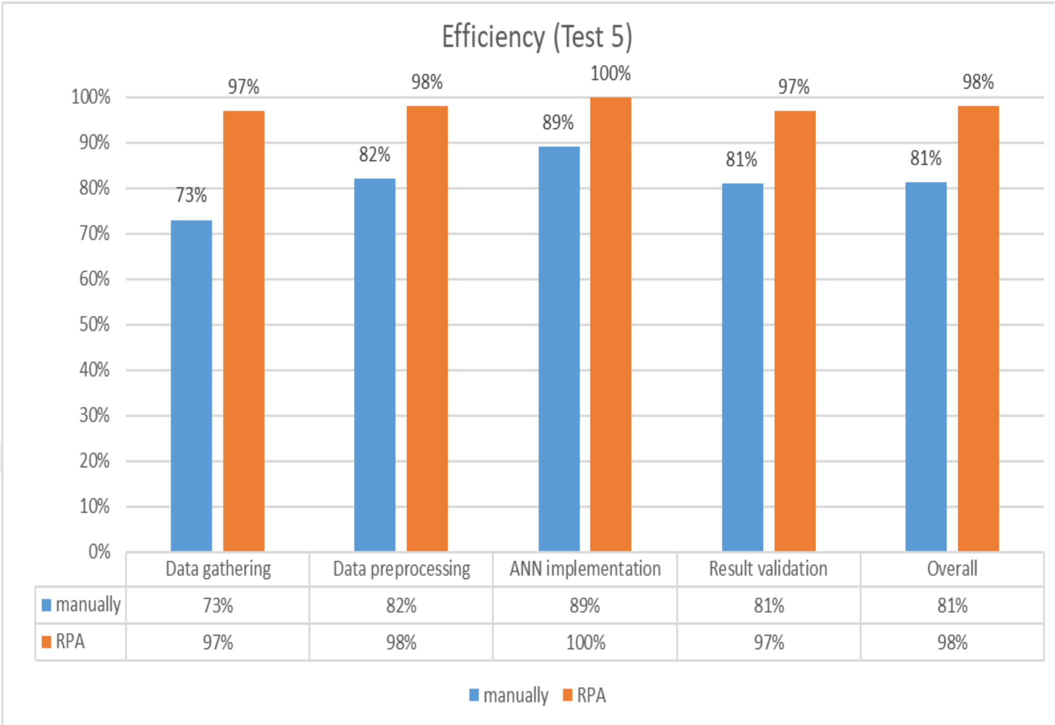



**Figure 5.10:** Comparison of Efficiency (5th test).

# 6. CHALLENGES

Even while RPA has increased organizational efficiency and enhanced employee performance, there are always issues that must be resolved as with any technology. We'll talk about some of the most typical RPA difficulties in this part, along with solutions.

i. Lack of Understanding: One of the major obstacles to robotic process automation is a lack of knowledge about how it operates and what it is capable of. When firms are contemplating whether or not to deploy RPA, this might be a problem. For assistance with this, a number of tools are available, including our own RPA 101 guide. Everything from what RPA is to how it can help your business is covered in this tutorial. [29]

ii. Implementation Costs: The price of putting RPA into practice is another frequent issue. RPA can be expensive up front, but it's vital to keep in mind that it can yield a return on investment (ROI) in a short amount of time. RPA costs can be decreased in a number of ways, including by utilizing open-source software and capitalizing on already existing infrastructure [29].

iii. Monitoring: it is an important part of RPA. Making sure your RPA solution is monitored is essential so you can identify and resolve any issues as soon as they appear. In order to keep you in control and provide you a complete picture of what the robots are doing, well-known RPA technologies include a central place to monitor robot performance and an execution plan. [29].

iv. Set moderate expectations at the outset: A viewpoint that is too enthusiastic is easy to embrace given the RPA hype. But it's crucial to keep your feet on the ground since doing otherwise has a big impact on how automation achievements are evaluated and, as a result, how decisions about scaling up to enterprise level are made. Determine how RPA may directly help in attaining each company aim after creating a clear hierarchy of those goals. [30]

v. Inability to fully automate operations: RPA systems might not be able to fully automate all the steps of more complicated procedures. Our suggested strategy is to "divide and conquer." Start here by redesigning these complex tasks, decomposing them into simpler components, and automating them [30].

vi. Overcoming RPA Challenges: RPA's automation features are perfect for repetitive, rule-based, high-volume processes that don't require human judgment. This can include operations like copy-and-paste work and data migration. However, non-standardized business processes that frequently require human intervention to complete make RPA installation particularly challenging [31].

vii. Final Thoughts: The healthcare business is being dramatically revolutionized day by day by robotic process automation, AI, and machine learning. RPA systems in the healthcare industry improve patient care while streamlining practice process [31]. These difficulties can be further divided into organizational, implementation, technical, and post-deployment problems.

viii. A lack of business and IT alignment: The business division might emphasize that RPA initiatives aim to relieve the workload of IT workers rather than entirely replace the system. Be ready to solve any security issues the IT department may have as well. You can create an RPA competency center with representatives from the business and IT to enable mutual understanding and open communication. [32]

A lack of ownership and poorly defined responsibilities: It is preferable to clearly state who is in charge of each project component, such as approving the design, monitoring the execution, and calculating the success rate.

ix. Lack of a clear RPA strategy. Every firm needs to create a plan before using RPA. During that time [32].

x. Selecting the incorrect procedure.

a. Process structure: seek out procedures that are template-driven, based on predefined standards, and necessitate repetitive manual input.

b. Change frequency: It will be difficult to automate a process that changes frequently.

c. Process complexity: A process is not a good candidate for automation if it calls for complicated cognitive functions. [32]

d. Business impact: Choose processes with a strong business impact to achieve immediate returns from your automation endeavour.

xi. Failing to streamline and optimize the chosen process before automating. To overcome this RPA difficulty, you can streamline and optimize a process by identifying and eliminating any extra stages, such as duplicative approvals, and, if possible, reducing the number of exceptions. [32]

xii. Lack of a sufficient infrastructure: RPA deployment will not be successful without a suitable infrastructure. Consider the capabilities you'll need and whether the infrastructure you have now can support all of this as you try to solve the RPA dilemma. You are seeking two things in your infrastructure in particular:

xiii. It must function continuously and be powerful enough to run all of your programs. This is where deploying a failover server will come in handy [32].

xiv. Scaling challenges. To lessen the hassle of scaling, implement your processes so that they operate independently and concurrently. Then, you can scale more easily than by raising the capability of a single bot by increasing the number of bots. Additionally, make sure the architecture of your RPA systems is streamlined and that they are audited after implementation [32].

xv. Security: Give each bot its own set of identification credentials. The implementation of two-factor identification for both operators would also be beneficial.

a. RPA access privileges should only be granted to the systems that each bot requires to do its duties. A bot should only be allowed read-only access if it reads data from a database. You may keep an eye on how bots interact with your systems by using screenshots or video monitoring.

b. Create RPA tools that can produce consistent, gap-free logs that can be checked for suspicious activity [32].

# 7. CONCLUSION AND FUTURE WORK

In this work, we presented an approach to exploiting RPA abilities in the investigation process of academic research work (in particular ANN research). This was carried out by presenting a case study of using the UiPath toolkit of RPA in the processes of conducting research investigation of predicting students' performance datasets using ANN method (PNN). In this approach model, we presented the potential steps in ANN-based research studies that can be automated using RPA toolkits. The employment of RPA, it is determined, can be a very helpful instrument for enhancing research duties and activities in the ANN research field and opens the door for a wide range of applications. Hence, ANN-based research requires different activities in data gathering, data mining and filtering, which require time and effort, and need to be performed as precisely as possible. The results show that the sequential steps of ANN can be easily automated using the proposed model and improve time consumption and performance quality. For future work, we intend to test the ability of the proposed model to be applied in more complex machine learning research such as image recognition and used tools that include coding tasks such MATLAB programming language. The results will be of interest to the research community of the ANN investigation.

# REFERENCES


[1] W. M. P. v. d. Aalst, M. Bichler and A. Heinzl, Robotic process automation, Springer, 2018.

[2] Infopédia, Dicionário Infopédia da Língua Portuguesa.

[3] R. Syed, S. Suriadi, M. Adams, W. Bandara, S. J. Leemans, C. Ouyang, A. H. t. Hofstede, I. v. d. Weerd and M. Th, “Robotic Process Automation: Contemporary themes and challenges,” Elsevier, p. 15, 2020.

[4] J. Siderska, “Robotic Process Automation—a driver of digital transformation?,” Engineering Management in Production and Services, vol. 12, 2020.

[5] B. C. Csáji, “Approximation with artificial neural networks,” Faculty of Sciences, Etvs Lornd University, Hungary, vol. 24, p. 7, 2001.

[6] Delloite, “Automation with intelligence Reimagining the organisation in the ‘Age of With’,”2019.[Online].Available: https://www2.deloitte.com/content/dam/Deloitte/tw/Documents/strategy/tw-Automation-with-intelligence.pdf. [Accessed 10 April 2021].

[7] F. Huang and M. A. Vasarhelyi, “Applying robotic process automation (RPA) in auditing: A framework,” International Journal of Accounting Information, vol. 35, p. 11, 2019.

[8] “Kaggle,” [Online]. Available: https://www.kaggle.com/spscientist/students-performance-in-exams/tasks. [Accessed 2 10 2021].

[9] E. Parzen, “On Estimation of a Probability Density Function and Mode,” Ann. Math. Statist., 1962.

[10] uipath, “What is robotic process automation?,” uipath, [Online]. Available: https://www.uipath.com/rpa/robotic-process-automation. [Accessed 15 August 2022].

[11] 1. M. P. R. P. A. R. T. I. 2022, "Software Testing Help," 10 Most Popular Robotic Process Automation RPA Tools In 2022, [Online].

Available: https://www.softwaretestinghelp.com/robotic-process-automation-tools/. [Accessed 7 August 2022].

[12] TechTarget, "What is robotic process automation?," TechTarget, [Online]. Available: https://www.techtarget.com/searchcio/definition/RPA. [Accessed 2022].

[13] uipath, "What is robotic process automation?," uipath, [Online]. Available: https://www.uipath.com/rpa/robotic-process-automation. [Accessed 15 August 2022].

[14] uipath, "What is robotic process automation?," uipath, [Online]. Available: https://www.uipath.com/rpa/robotic-process-automation. [Accessed 15 August 2022].

[15] uipath, "What is robotic process automation?," uipath, [Online]. Available: https://www.uipath.com/rpa/robotic-process-automation. [Accessed 15 august 2022].

[16] Uipath, "What is robotic process automation?," Uipath, [Online]. Available: https://www.uipath.com/rpa/robotic-process-automation. [Accessed 15 August 2022].

[17] ei-path, "Robotic Process Automation (RPA)," ei-path, [Online]. Available: https://ei-path.com/service/process-automation-using-rpa/. [Accessed 2021].

[18] A. Anywhere, "What is Macro Recording?," Automation Anywhere, [Online]. Available: https://www.automationanywhere.com/rpa/macro-recording. [Accessed What is Macro Recording?].

[19] B. M. GmbH, "Record mouse and keyboard actions for infinite replay," Bartels Media GmbH, [Online]. Available: https://www.macrorecorder.com. [Accessed 2022].

[20] W. i. m. learning?, "What is machine learning?," What is machine learning?, [Online]. Available: https://www.ibm.com/cloud/learn/machine-learning. [Accessed 15 July 2020].

[21] TechTarget, "machine learning," TechTarget, [Online]. Available: https://www.techtarget.com/searchenterpriseai/definition/machine-learning-

ML#:~:text=Machine%20learning%20is%20important%20because,central%20part%20of%20their%20operations.. [Accessed 2022].

[22] E. Burns, "machine learning," techtarget, [Online]. Available: https://www.techtarget.com/searchenterpriseai/definition/machine-learning-ML. [Accessed 2022].

[23] E. Burns, "machine learning," TechTarget, [Online]. Available: https://www.techtarget.com/searchenterpriseai/definition/machine-learning-ML. [Accessed 2022].

[24] M. Blog, "What are the advantages and disadvantages of machine learning?," My Blog, [Online]. Available: https://ai.foobrdigital.com/what-are-the-advantages-and-disadvantages-of-machine-learning/. [Accessed 2022].

[25] J. Selig, "What Is Machine Learning? A Definition," EXPERT.AI, [Online]. Available: https://www.expert.ai/blog/machine-learning-definition/. [Accessed 14 march 2022].

[26] V. Kanade, "What Is Machine Learning? Definition, Types, Applications, and Trends for 2022," Spiceworks Inc., [Online]. Available: https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-ml/. [Accessed 30 august 2022 ].

[27] I. C. Education, " What is Machine Learning?," IBM Cloud Education, [Online]. Available: https://www.ibm.com/cloud/learn/machine-learning. [Accessed 15 july 2020].

[28] S. Dwivedi, P. Kasliwal and S. Soni, "Comprehensive study of data analytics tools (RapidMiner, Weka, R tool, Knime)," 2016 Symposium on Colossal Data Analysis and Networking (CDAN), 2016.

[29] H. A. MALAK, "9 Robotic Process Automation Challenges & How to Solve Them," The ECM Consultant, [Online]. Available: https://theecmconsultant.com/robotic-process-automation-challenges/. [Accessed 22 AUGUST 2022].

[30] CiGen, “7 Challenges to Implementing Robotic Process Automation (RPA) & How to Overcome Them,” CiGen , [Online]. Available: https://www.cigen.com.au/7-challenges-implementing-rpa-how-overcome-them/. [Accessed 20 January 2020].

[31] capminds, “The Top Challenges Of RPA Implementation,” capminds, [Online]. Available: https://www.capminds.com/blog/the-top-challenges-of-rpa-implementation/. [Accessed 5 February ,2021].

[32] I. A. Nadejda Alkhaldi, “Top 13 RPA challenges and ways to overcome them,” Itrex Group, [Online]. Available: https://itrexgroup.com/blog/top-rpa-challenges-and-ways-to-overcome-them/. [Accessed 8 September 2022].

[33] U. S. introduction, “ UiPath,” UiPath, 2020. [Online]. Available: https://docs.uipath.com/studio/docs/introduction. [Accessed 10 April 2021].

# APPENDIX A

A.1 Sample of Unilu Gaia trace:

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Job Number | Submit Time | Wait Time | Run Time | Number of Allocated Processors | Average CPU time used | Used Memory | Requested Number of Processors | Requested Time | Requested Memory | Status | User ID | Group ID | Executable Number | Queue Number | Partition Number | Preceding Job Number | Think Time from Preceding Job |
| 2 | 6432 | 10411388 | 1 | 284 | 1 | -1 | -1 | 1 | 7200 | -1 | 1 | 37 | 37 | -1 | 1 | 1 | -1 | -1 |
| 3 | 39067 | 28540932 | 6 | 62 | 20 | -1 | -1 | 20 | 7200 | -1 | 1 | 57 | 57 | -1 | 1 | 1 | -1 | -1 |
| 4 | 44458 | 30150222 | 0 | 2915 | 40 | -1 | -1 | 40 | 7200 | -1 | 1 | 45 | 45 | -1 | 1 | 1 | -1 | -1 |
| 5 | 3883 | 6947060 | 1 | 930 | 20 | -1 | -1 | 20 | 18000 | -1 | 1 | 17 | 17 | -1 | 1 | 1 | -1 | -1 |
| 6 | 48996 | 32684475 | 4 | 3422 | 20 | -1 | -1 | 20 | 259200 | -1 | 1 | 46 | 46 | -1 | 1 | 1 | -1 | -1 |
| 7 | 27687 | 26049051 | 8 | 318 | 20 | -1 | -1 | 20 | 259200 | -1 | 1 | 64 | 64 | -1 | 1 | 1 | -1 | -1 |
| 8 | 25979 | 24923637 | 0 | 63 | 100 | -1 | -1 | 100 | 5400 | -1 | 1 | 17 | 17 | -1 | 1 | 1 | -1 | -1 |
| 9 | 25146 | 24329023 | 4 | 326 | 20 | -1 | -1 | 20 | 3600 | -1 | 1 | 42 | 42 | -1 | 1 | 1 | -1 | -1 |
| 10 | 32442 | 27379220 | 4 | 321 | 20 | -1 | -1 | 20 | 1200 | -1 | 1 | 57 | 57 | -1 | 1 | 1 | -1 | -1 |
| 11 | 7875 | 13169402 | 3 | 133 | 80 | -1 | -1 | 80 | 3600 | -1 | 1 | 6 | 6 | -1 | 1 | 1 | -1 | -1 |
| 12 | 35976 | 28309109 | 1899 | 386 | 20 | -1 | -1 | 20 | 2700 | -1 | 1 | 57 | 57 | -1 | 1 | 1 | -1 | -1 |
| 13 | 49637 | 32738919 | 4039 | 69 | 20 | -1 | -1 | 20 | 9000 | -1 | 1 | 2 | 2 | -1 | 1 | 1 | -1 | -1 |
| 14 | 14862 | 19316699 | 129 | 108 | 5120 | -1 | -1 | 5120 | 1200 | -1 | 1 | 49 | 49 | -1 | 1 | 1 | -1 | -1 |
| 15 | 2731 | 4107707 | 1 | 169765 | 580 | -1 | -1 | 580 | 216000 | -1 | 1 | 20 | 20 | -1 | 1 | 1 | -1 | -1 |
| 16 | 35932 | 28309093 | 1579 | 378 | 20 | -1 | -1 | 20 | 2700 | -1 | 1 | 57 | 57 | -1 | 1 | 1 | -1 | -1 |
| 17 | 5113 | 8975734 | 0 | 119 | 20 | -1 | -1 | 20 | 600 | -1 | 1 | 33 | 33 | -1 | 1 | 1 | -1 | -1 |
| 18 | 36582 | 28374996 | 6 | 74 | 20 | -1 | -1 | 20 | 2700 | -1 | 1 | 57 | 57 | -1 | 1 | 1 | -1 | -1 |
| 19 | 47392 | 31670491 | 33 | 44963 | 40 | -1 | -1 | 40 | 259200 | -1 | 1 | 103 | 103 | -1 | 1 | 1 | -1 | -1 |
| 20 | 18192 | 21306122 | 10 | 47 | 20 | -1 | -1 | 20 | 300 | -1 | 1 | 6 | 6 | -1 | 1 | 1 | -1 | -1 |
| 21 | 45411 | 30462558 | 1 | 151 | 320 | -1 | -1 | 320 | 600 | -1 | 1 | 33 | 33 | -1 | 1 | 1 | -1 | -1 |
| 22 | 21794 | 22765661 | 5 | 600 | 1 | -1 | -1 | 1 | 600 | -1 | 1 | 40 | 40 | -1 | 1 | 1 | -1 | -1 |
| 23 | 33227 | 27462325 | 3 | 454 | 20 | -1 | -1 | 20 | 600 | -1 | 1 | 57 | 57 | -1 | 1 | 1 | -1 | -1 |

Sample of RICC trace:

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Job Number | Submit Time | Wait Time | Run Time | Number of Allocated Processors | Average CPU time used | Used Memory | Requested Number of Processors | Requested Time | Requested Memory | Status | User ID | Group ID | Executable Number | Queue Number | Partition Number | Preceding Job Number |
| 2 | 2591 | 1360775 | 6 | 1240 | 128 | 423 | 3108128 | 128 | -1 | -1 | 0 | 61 | 1 | 1 | 16 | -1 | -1 |
| 3 | 5168685 | 39360 | 23648 | 128 | 18208 | 8447616 | 128 | -1 | -1 | 1 | 23 | 1 | 1 | 3 | -1 | -1 | -1 |
| 4 | 11061568 | 6 | 15 | 4 | -1 | 4336 | 4 | -1 | -1 | 1 | 51 | 1 | 1 | 14 | -1 | -1 | -1 |
| 5 | 2331227 | 6 | 1435 | 1 | 1244 | 476752 | 1 | -1 | -1 | 1 | 12 | 1 | 1 | 11 | -1 | -1 | -1 |
| 6 | 8689535 | 79 | 55 | 128 | -1 | -1 | 128 | -1 | -1 | 1 | 61 | 1 | 1 | 3 | -1 | -1 | -1 |
| 7 | 4488380 | 16 | 334 | 2 | 150 | 14992 | 2 | -1 | -1 | 0 | 212 | 1 | 1 | 14 | -1 | -1 | -1 |
| 8 | 4382492 | 36277 | 96 | 1 | 12 | 12096 | 1 | -1 | -1 | 1 | 37 | 1 | 1 | 3 | -1 | -1 | -1 |
| 9 | 9767 | 2147885 | 15 | 755 | 1 | 493 | 1841296 | 1 | -1 | -1 | 1 | 58 | 1 | 136 | 3 | -1 | -1 |
| 10 | 2763705 | 5 | 268 | 8 | 2 | 113072 | 8 | -1 | -1 | 0 | 147 | 1 | 1097 | 11 | -1 | -1 | -1 |
| 11 | 4967102 | 487 | 3405 | 128 | 2550 | 3140896 | 128 | -1 | -1 | 1 | 61 | 1 | 1 | 16 | -1 | -1 | -1 |
| 12 | 3883043 | 6 | 361 | 1 | 355 | 18816 | 1 | -1 | -1 | 1 | 143 | 1 | 1 | 3 | -1 | -1 | -1 |
| 13 | 7552037 | 5 | 1438 | 48 | 641 | 497184 | 48 | -1 | -1 | 1 | 13 | 1 | 8 | 16 | -1 | -1 | -1 |
| 14 | 9244256 | 5 | 6678 | 1 | 1 | 11600 | 1 | -1 | -1 | 0 | 79 | 1 | 1 | 14 | -1 | -1 | -1 |
| 15 | 2272689 | 6 | 1928 | 64 | 1170 | 717392 | 64 | -1 | -1 | 1 | 13 | 1 | 7 | 16 | -1 | -1 | -1 |
| 16 | 9220079 | 406 | 52 | 1 | 4 | 5024 | 1 | -1 | -1 | 1 | 11 | 1 | 2488 | 14 | -1 | -1 | -1 |
| 17 | 2658627 | 6 | 28755 | 1 | 273 | 13216 | 1 | -1 | -1 | 1 | 30 | 1 | 1 | 14 | -1 | -1 | -1 |
| 18 | 6789797 | 294 | 9 | 129 | -1 | 2736 | 129 | -1 | -1 | 0 | 77 | 1 | 2047 | 8 | -1 | -1 | -1 |
| 19 | 11731688 | 7 | 190 | 12 | -1 | 2464 | 12 | -1 | -1 | 1 | 5 | 1 | 1 | 16 | -1 | -1 | -1 |
| 20 | 5191338 | 572770 | 47 | 128 | -1 | 2576 | 128 | -1 | -1 | 1 | 61 | 1 | 1 | 3 | -1 | -1 | -1 |
| 21 | 8670683 | 11 | 3564 | 64 | 2616 | 698048 | 64 | -1 | -1 | 1 | 13 | 1 | 2254 | 16 | -1 | -1 | -1 |
| 22 | 8745463 | 11 | 1933 | 48 | 940 | 496336 | 48 | -1 | -1 | 1 | 13 | 1 | 2196 | 16 | -1 | -1 | -1 |
| 23 | 8567864 | 5 | 5 | 1 | 1 | 2800 | 1 | -1 | -1 | 1 | 5 | 1 | 1 | 16 | -1 | -1 | -1 |

Sample of ANL-interpad trace:

| | A | B | C | D | E | F | G | H | I | J | K | L | M |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Job Number | Submit Time | Wait Time | Run Time | Number of Allocated Processors | Average CPU time used | Used Memory | Requested Number of Processors | Requested Time | Requested Memory | Status | User ID | Group ID |
| 2 | 7001 | 6298047 | -1 | 110 | 64 | -1 | -1 | 64 | -1 | -1 | 0 | 18 | -1 |
| 3 | 15028865 | -1 | 449 | 512 | -1 | -1 | 512 | 1800 | -1 | 0 | 96 | -1 | 3221 |
| 4 | 9331 | 7534176 | -1 | 0 | -1 | -1 | -1 | -1 | -1 | -1 | 5 | 20 | -1 |
| 5 | 16139706 | -1 | 10 | 8 | -1 | -1 | 8 | 1800 | -1 | 5 | 94 | -1 | 2927 |
| 6 | 7773 | 6568782 | -1 | 1 | 48 | -1 | -1 | 48 | -1 | -1 | 1 | 24 | -1 |
| 7 | 18393054 | -1 | 1823 | 128 | -1 | -1 | 128 | 1800 | -1 | 0 | 20 | -1 | 1636 |
| 8 | 18727311 | -1 | 0 | -1 | -1 | -1 | -1 | 864000 | -1 | 5 | 116 | -1 | 4203 |
| 9 | 15743749 | -1 | 9 | 8 | -1 | -1 | 8 | 7200 | -1 | 5 | 82 | -1 | 2849 |
| 10 | 9188467 | -1 | 6057 | 64 | -1 | -1 | 64 | 7200 | -1 | 1 | 52 | -1 | 1654 |
| 11 | 9420 | 7590203 | -1 | 13 | 16 | -1 | -1 | 16 | 3600 | -1 | 0 | 4 | -1 |
| 12 | 8704 | 7156623 | -1 | 38 | 512 | -1 | -1 | 512 | -1 | -1 | 5 | 36 | -1 |
| 13 | 2781 | 4744637 | -1 | 640 | 64 | -1 | -1 | 64 | 3600 | -1 | 1 | 4 | -1 |
| 14 | 15117263 | -1 | 11 | 128 | -1 | -1 | 128 | 1800 | -1 | 0 | 70 | -1 | 3253 |
| 15 | 8112175 | -1 | 1 | 8 | -1 | -1 | 8 | -1 | -1 | 1 | 2 | -1 | 186 |
| 16 | 7318 | 6448398 | -1 | 7 | 8 | -1 | -1 | 8 | -1 | -1 | 1 | 19 | -1 |
| 17 | 18817074 | -1 | 75 | 128 | -1 | -1 | 128 | 7200 | -1 | 1 | 110 | -1 | 4229 |
| 18 | 19738080 | -1 | 1765 | 1024 | -1 | -1 | 1024 | 7200 | -1 | 1 | 103 | -1 | 4163 |
| 19 | 16851263 | -1 | 536 | 1024 | -1 | -1 | 1024 | 7200 | -1 | 1 | 103 | -1 | 3599 |
| 20 | 2396 | 4270784 | -1 | 385 | 288 | -1 | -1 | 288 | 10800 | -1 | 1 | 3 | -1 |
| 21 | 9065 | 7474323 | -1 | 200 | 64 | -1 | -1 | 64 | -1 | -1 | 1 | 37 | -1 |

Sample of METACENTRUM trace:

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | job ID | Submit time | Wait Time | RunTime | N. alloc Proc | Average time cpu | Used Mem | Req No. proc | Req Time | Req Mem | Status | User ID | Group ID | Executabl e | Queue | Partition | Job N. | Think time | Column1. 19 |
| 2 | 15959 | 8359611 | 0 | 7 | 5 | 6 | 0.48 | -1 | -1 | -1 | -1 | 1 | 50 | -1 | -1 | 0 | 1 | -1 | -1 |
| 3 | 40534 | 16365006 | 0 | 355 | 1 | 367 | 0.2 | -1 | -1 | -1 | -1 | 1 | 57 | -1 | -1 | 0 | 1 | -1 | -1 |
| 4 | 56041 | 26031896 | 16220 | 269 | 32 | 202 | 0.16 | -1 | -1 | -1 | -1 | 1 | 31 | -1 | -1 | 8 | 2 | -1 | -1 |
| 5 | 1609 | 1006472 | 11 | 26604 | 8 | 26562 | -1 | -1 | -1 | -1 | 1 | 8 | -1 | -1 | 20 | 2 | -1 | -1 | |
| 6 | 43336 | 17858355 | 0 | 27 | 8 | 27 | 0.9 | -1 | -1 | -1 | -1 | 1 | 13 | -1 | -1 | 0 | 1 | -1 | -1 |
| 7 | 69589 | 29225461 | 0 | 14 | 8 | 13 | 0.95 | -1 | -1 | -1 | -1 | 1 | 32 | -1 | -1 | 0 | 1 | -1 | -1 |
| 8 | 67902 | 29059959 | 0 | 2 | 4 | 1 | 0.8 | -1 | -1 | -1 | -1 | 1 | 18 | -1 | -1 | 0 | 1 | -1 | -1 |
| 9 | 1908 | 1085257 | 16 | 32540 | 8 | 32506 | -1 | -1 | -1 | -1 | 1 | 8 | -1 | -1 | 20 | 2 | -1 | -1 | |
| 10 | 2316 | 1231099 | 0 | 22 | 8 | 117 | 0 | -1 | -1 | -1 | -1 | 1 | 5 | -1 | -1 | 0 | 1 | -1 | -1 |
| 11 | 13212 | 7002791 | 0 | 20 | 2 | 19 | 0.8 | -1 | -1 | -1 | -1 | 1 | 35 | -1 | -1 | 0 | 1 | -1 | -1 |
| 12 | 6088 | 3019533 | 148614 | 43375 | 32 | 43322 | -1 | -1 | -1 | -1 | 1 | 7 | -1 | -1 | 21 | 2 | -1 | -1 | |
| 13 | 31716 | 12878252 | 0 | 1 | 1 | -1 | -1 | -1 | -1 | -1 | 1 | 66 | -1 | -1 | 0 | 1 | -1 | -1 | |
| 14 | 14317 | 7425323 | 5 | 6 | 4 | -1 | -1 | -1 | -1 | -1 | 1 | 20 | -1 | -1 | 2 | 2 | -1 | -1 | |
| 15 | 45216 | 19095373 | 5 | 9153 | 31 | 9140 | -1 | -1 | -1 | -1 | 1 | 27 | -1 | -1 | 11 | 2 | -1 | -1 | |
| 16 | 28953 | 11872778 | 485 | 206 | 16 | 180 | 0.68 | -1 | -1 | -1 | -1 | 1 | 20 | -1 | -1 | 20 | 2 | -1 | -1 |
| 17 | 1529 | 932689 | 0 | 7 | 1 | 14 | 0.4 | -1 | -1 | -1 | -1 | 1 | 11 | -1 | -1 | 0 | 1 | -1 | -1 |
| 18 | 10997 | 5904821 | 0 | 177 | 2 | 176 | 0.4 | -1 | -1 | -1 | -1 | 1 | 35 | -1 | -1 | 0 | 1 | -1 | -1 |
| 19 | 14655 | 7636324 | 0 | 2 | 4 | 1 | 0.8 | -1 | -1 | -1 | -1 | 1 | 21 | -1 | -1 | 0 | 1 | -1 | -1 |
| 20 | 41449 | 16735384 | 0 | 31 | 4 | 30 | 0.6 | -1 | -1 | -1 | -1 | 1 | 34 | -1 | -1 | 0 | 1 | -1 | -1 |
| 21 | 63801 | 27218540 | 0 | 14 | 4 | 13 | 0.5 | -1 | -1 | -1 | -1 | 1 | 92 | -1 | -1 | 0 | 1 | -1 | -1 |
| 22 | 38995 | 15417166 | 8406 | 514 | 64 | 484 | 0.43 | -1 | -1 | -1 | -1 | 1 | 18 | -1 | -1 | 22 | 2 | -1 | -1 |
| 23 | 45901 | 19729649 | 5 | 16712 | 16 | 16690 | -1 | -1 | -1 | -1 | 1 | 3 | -1 | -1 | 20 | 2 | -1 | -1 | |
| 24 | 29347 | 12011584 | 0 | 108 | 16 | 107 | 0.77 | -1 | -1 | -1 | -1 | 1 | 25 | -1 | -1 | 0 | 1 | -1 | -1 |