

# Robotic Process Automation with Increasing Productivity using Machine Learning

Ayush Mishra<sup>1</sup>, Sauraw Diwase<sup>2</sup>, Rutik Kulkarni<sup>3</sup>, Govind Bhujbal<sup>4</sup>

Students, Department of Computer Engineering<sup>1,2,3,4</sup>  
Sinhgad Institute of Technology, Lonavala, Maharashtra, India

**Abstract:** For any organizations or government, public/private sector who is involve in purchasing & supply chain management for that organizations it is important to know the market trend for particular product sells. As we know that big organizations buy or produce any product in large nubers but to know the trend of that product is very important before buying or producing because it cost a lot for that organization. Organizations profit & loss is directly connected with the market trend. To solve this problem we have come up with a solution which involves integration of three technologies named Robotic Process Automation (RPA), Machine Learning (ML), Power BI. This paper explain how we can solve the issue of purchasing & supply chain management, analysing trend of particular product sells, making right business decisions. Our aim will be to develop a system which will accurately predict the market trend of any product sells.

**Keywords:** Robotic process automation, Purchasing and supply management (PSM), Logistic Regression, Random Forest Classifier, Power BI

## I. INTRODUCTION

RPA is defined as an art of using software robots to interact with Software-as-a-Service applications and IT systems to automate the rule-based manual jobs associated with repetitive and transactional processes.

By using RPA we will get huge amount of data in DB from any website and by using ML algorithm we will increases its productivity. However, the adoption of RPA and other advanced digital technologies in the PSM function impacts its processes, capabilities, and professionals.

Microsoft Power BI is used to find insights within an organization's data. Power BI can help connect disparate data sets, transform and clean the data into a data model and create charts or graphs to provide visuals of the data. All of this can be shared with other Power BI users within the organization.

## II. LITERATURE SURVEY

In the study of IEEE paper published by Devansh Hiren Timbadia, Parin Jigishu Shah, Sugosh Sudhanvan, Supriya Agrawal in their paper [1] titled “ Robotic Process Automation Through Advance Process Analysis Model ” Integrated Definition Group of Modelling Method & Artificial Intelligence Method have used to decide whether given task should be automated or not. In this paper author are telling us the saving one can do in company via using RPA. They have categorised task based on complexity of task and full time saving based on these feature they explained which task to automate & which task should not be automated.

IEEE paper by Ranjitha P, Spandana M, Amrita School of Arts and Sciences, Mysuru Amrita Vishwa Vidyapeetham, India titled “Predictive Analysis for Big Mart Sales Using Machine Learning Algorithms” [2] have worked on various algorithms such as Linear Regression,

Polynomial Regression, Ridge Regression, XgBoost Regression for better prediction of sells. The comparision factor which they have used are Accuracy, RMSE, MAE, MSE after all observation they conclude that ridge and Xgboost regression gives the better prediction with respect to Accuracy, MAE and RMSE than the Linear and polynomial regression. The authors have focused much more on accuracy.

Md. Anisur Rahman Mia, Mohammad Abu Yousuf in their paper titled “Business Forecasting System using Machine Learning Approach” [3] mainly focus on designing forecasting system using Backpropagation Neural Network (BPNN), Feed Forward Neural Network (FFNN), Deep Neural Network (DNN)

To implement the designed system to forecast sales and analyse result. Author of this paper have drastically reduced the error using BPNN. Iterative Back Propagation took 10000 times to make training error minimum. Overall accuracy is between 75% to 99% on test data set.

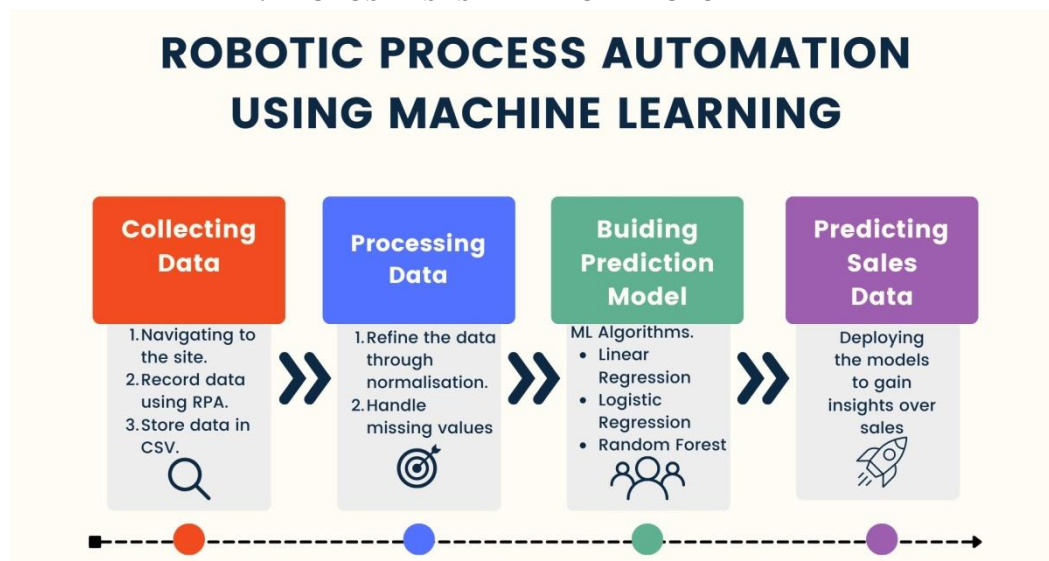
Mohammed Ameer, Simhadri Prem Rahul, Dr.Suneetha Manne in their research paper titled “Human Resource Analytics using Power Bi Visualization Tool” [4] focuses on Human Resource management by predictive analysis on employee data such as employee turnover analysis, employee work performance analysis, and training requirements analysis. Author uses Power BI tool on real-time data. With the help of Power BI , Logistic Regression, Gini – index author make a model which can boost up the economic growth of company.

**Table I:** Comparison Between Literature Papers

Features	[1]	[2]	[3]	[4]
Purpose	After analysing various parameters, it decides given task is suitable for automation or not	Better prediction for developing & improving market strategies	to predict its future sale, profit and loss.	To boost up the economic growth of a company as it uses simple dashboards for performance levels.
Algorithm used	Integrated Definition Group of Modelling Method, Artificial Intelligence Method	Linear Regression, Polynomial Regression, Ridge Regression, XgBoost Regression	Back propagation neural network (BPNN) Deep Learning Network (DNN), Feed Forward Neural Network ( FFNN )	Logistic Regression, Random Forest Classifier, Power BI (Visualisation Tools )
Focus on	classify the process and provide suitability of the process for RPA	The Method of forecasting for predicting sells value is very beneficial towards ERP system.	ML technology is implemented to develop a prediction system and to forecast sale using the system. ML finds specific patterns or relationship from the training data.	study aims to increase to identify employee churn and demand using POWER BI with the help of real-time data insights such as dashboards which run machine learning models like Logistic Regression and Random Forest in background
Scope of improvement	factors like cost estimation, paybacks, and benefits	Unforeseen cash flow and unmanaged production, staff and financing.	Training error was found maximum for the commodity that has much price deviation over time.	We have to use CNN & RNN when there is huge amount of dataset.

Conclusion	RPA is centered solution for optimizing human usage on mundane repetitive boring tasks	ridge and Xgboost regression gives the better prediction with respect to Accuracy, MAE and RMSE than the Linear and polynomial regression approaches.	designed system is very efficient to predict sales as overall prediction accuracy varied from 99% to 75% for test data of different product categories	It can be observed that the evaluation scores of random forest are better than logistic regression and so is the reason random forest is considered for the model training. So results of Random Forest classifier are better than those of logistic regression and hence the model that is trained by the random forest classifier acts as insights to POWER BI
------------	--	---	--	--

### III. PROPOSED SYSTEM ARCHITECTURE



### IV. EXPERIMENTAL SETUP

#### Software Tools Used

- Windows 10 operating system
- Browser: Chrome/Firefox
- Automation A360
- Jupyter Notebook
- Microsoft Power BI Desktop
- Microsoft Office

The Dataset which we have used for this work can be accessed using link mentioned in double quotes “<https://drive.google.com/file/d/1HrPdFUtZia2Fk2HTIJvW-riNJYI-1Ak/view?usp=sharing>”

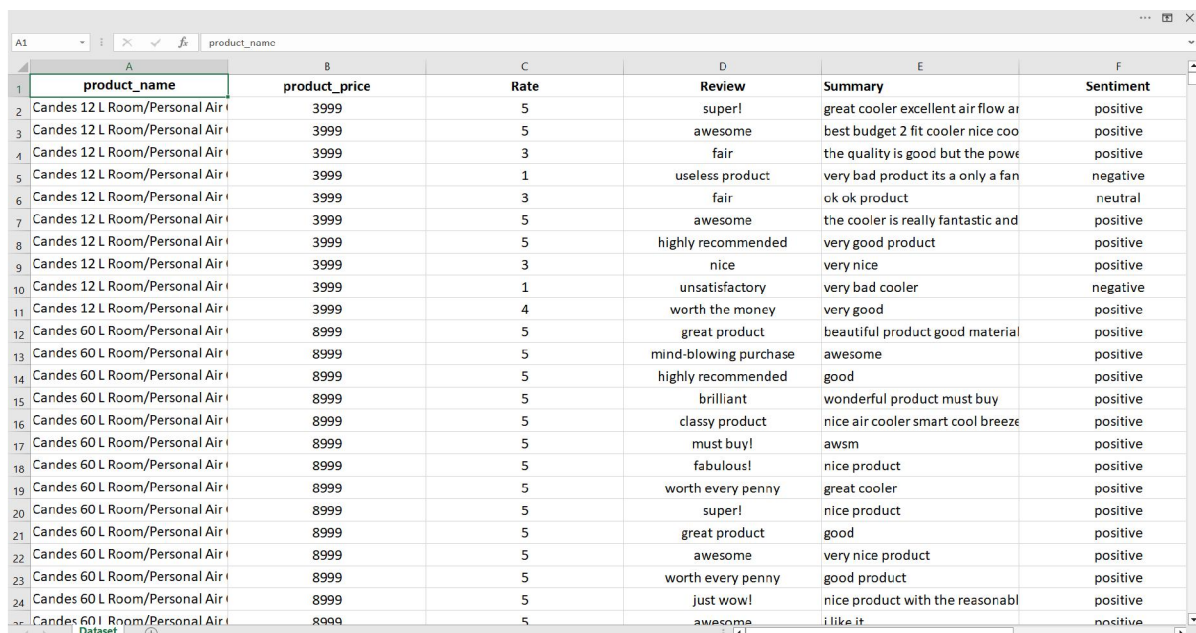
## V. METHODOLOGY

Fetching data from run-time website like flipkart using RPA

**Dataset:** We’ve collected the dataset from the internet for the website called kaggle.com .In this work all having test dataset and train dataset in the test data set having a 5000 dataset and in the train data having a 8000 data

We’ve created task bot using Automation 360 . We’ve deployed our created task bot on flipkart website/server.

As per the instruction we have given to that task bot it will do the same task for specified no of time eventually it will fetch the data from that website and store it in Microsoft Excel. Fig. 1 shows the data we have fetched using RPA



	A	B	C	D	E	F
	product_name	product_price	Rate	Review	Summary	Sentiment
1	Candes 12 L Room/Personal Air	3999	5	super!	great cooler excellent air flow ar	positive
2	Candes 12 L Room/Personal Air	3999	5	awesome	best budget 2 fit cooler nice coo	positive
3	Candes 12 L Room/Personal Air	3999	3	fair	the quality is good but the powe	positive
4	Candes 12 L Room/Personal Air	3999	1	useless product	very bad product its a only a fan	negative
5	Candes 12 L Room/Personal Air	3999	3	fair	ok ok product	neutral
6	Candes 12 L Room/Personal Air	3999	5	awesome	the cooler is really fantastic and	positive
7	Candes 12 L Room/Personal Air	3999	5	highly recommended	very good product	positive
8	Candes 12 L Room/Personal Air	3999	3	nice	very nice	positive
9	Candes 12 L Room/Personal Air	3999	1	unsatisfactory	very bad cooler	negative
10	Candes 12 L Room/Personal Air	3999	4	worth the money	very good	positive
11	Candes 60 L Room/Personal Air	8999	5	great product	beautiful product good material	positive
12	Candes 60 L Room/Personal Air	8999	5	mind-blowing purchase	awesome	positive
13	Candes 60 L Room/Personal Air	8999	5	highly recommended	good	positive
14	Candes 60 L Room/Personal Air	8999	5	brilliant	wonderful product must buy	positive
15	Candes 60 L Room/Personal Air	8999	5	classy product	nice air cooler smart cool breeze	positive
16	Candes 60 L Room/Personal Air	8999	5	must buy!	awsm	positive
17	Candes 60 L Room/Personal Air	8999	5	fabulous!	nice product	positive
18	Candes 60 L Room/Personal Air	8999	5	worth every penny	great cooler	positive
19	Candes 60 L Room/Personal Air	8999	5	super!	nice product	positive
20	Candes 60 L Room/Personal Air	8999	5	great product	good	positive
21	Candes 60 L Room/Personal Air	8999	5	awesome	very nice product	positive
22	Candes 60 L Room/Personal Air	8999	5	worth every penny	good product	positive
23	Candes 60 L Room/Personal Air	8999	5	just wow!	nice product with the reasonabl	positive
24	Candes 60 L Room/Personal Air	8999	5	awesome	i like it	positive

Fig. 1 fetched dataset from flipkart website

Applying Machine Learning Algorithms & Building a model

### Linear Regression

Linear regression is a popular machine learning algorithm used for predicting a continuous output variable based on one or more input features. In the case of sales prediction, we can use linear regression to predict the sales based on the price and rating columns.

1. Data Preprocessing: Preprocess the data by normalizing or standardizing the price and rating data to make it suitable for analysis.
2. Data Splitting: Split the data into training and testing datasets.
3. Feature Selection: Select the price and rating columns as features for the analysis.
4. Model Training: Train the linear regression model on the training data to predict the sales based on the price and rating features.
5. Model Evaluation: Evaluate the performance of the model on the testing dataset using appropriate evaluation metrics such as mean squared error (MSE), root mean squared error (RMSE), or R-squared.
6. Sales Prediction: Use the trained linear regression model to predict the sales for the new products based on their price and rating values.

### Logistic Regression :

Logistic regression is a machine learning algorithm used for predicting a binary output variable based on one or more input features. In the case of sales prediction, we can use logistic regression to predict the likelihood of a product being sold based on its price and rating.

1. Data Preprocessing: Preprocess the data by normalizing or standardizing the price and rating data to make it suitable for analysis.
2. Data Splitting: Split the data into training and testing datasets.
3. Feature Selection: Select the price and rating columns as features for the analysis.
4. Model Training: Train the logistic regression model on the training data to predict the likelihood of a product being sold based on the price and rating features.
5. Model Evaluation: Evaluate the performance of the model on the testing dataset using appropriate evaluation metrics such as accuracy, precision, recall, or F1-score.
6. Sales Prediction: Use the trained logistic regression model to predict the likelihood of a product being sold based on its price and rating values.

#### **Random forest :**

Random forest is a machine learning algorithm used for predicting a continuous or categorical output variable based on one or more input features. In the case of sales prediction, we can use random forest to predict the sales based on the price and rating columns.

1. Data Preprocessing: Preprocess the data by normalizing or standardizing the price and rating data to make it suitable for analysis.
2. Data Splitting: Split the data into training and testing datasets.
3. Feature Selection: Select the price and rating columns as features for the analysis.
4. Model Training: Train the random forest model on the training data to predict the sales based on the price and rating features.
5. Model Evaluation: Evaluate the performance of the model on the testing dataset using appropriate evaluation metrics such as mean squared error (MSE), root mean squared error (RMSE), or R-squared.
6. Sales Prediction: Use the trained random forest model to predict the sales for the new products based on their price and rating values. Overall, the process for using Linear Regression, Logistic Regression, and Random Forest for sales prediction using price and rating columns is similar. The key differences lie in the algorithm used for model training and the evaluation metrics used to assess the performance of the models.

#### **Visualizing the output using Power BI tool**

Power BI is a powerful business intelligence tool that can help visualize and analyze data. Robotic Process Automation (RPA) can be used to automate repetitive tasks in Power BI, such as data extraction and report generation. To visualize the output using Power BI and RPA, you can follow these steps: Connect to your data source: Power BI supports a wide range of data sources, including Excel, CSV, SQL Server, and many others. You can use RPA to extract data from your source and load it into Power BI. Create a report: Once you have connected to your data source, you can use Power BI to create a report. This can include charts, tables, and other visualizations that help you better understand your data. Automate report generation: RPA can be used to automate the process of generating reports in Power BI. This can save time and reduce errors. Schedule report delivery: You can use RPA to schedule the delivery of reports to stakeholders. This ensures that everyone has access to the latest data and insights. Overall, using Power BI and RPA together can help you gain valuable insights from your data, while automating repetitive tasks and saving time.

### **VI. APPLICATIONS**

1. Ease in Decision Making Process Robotic Process Automation (RPA) combined with machine learning (ML) applications can greatly enhance productivity and improve product quality in business operations [1]. By incorporating cognitive technologies such as ML, speech recognition, and natural language processing, higher-order tasks can be automated with AI assistance [1].
2. Ease in Purchasing & supply chain Management: The combination of RPA and ML can produce a fully automated solution for tasks such as document extraction and process automation [3]. While RPA is good for rote and repetitive tasks, ML is more dynamic and can adapt to changes, making it a powerful tool for intelligent automation [2].



3. Help business to analyse market needs To incorporate AI in process automation, it is important to have a fundamental understanding of both technologies, including historical data for model training, process understanding, actors, trigger points, sub-system boundaries, domain knowledge, interfacing API/hooks, regulations, and exception handling [1]. However, it is important to note that automation should not be applied to poorly designed business processes, as it will only accelerate inefficiencies [1].

## VII. CONCLUSION

In this work we have integrated RPA, ML & Power BI to analyse the past sells of product & predict the future sells of product.

Using RPA we can fetch the various Data from various websites. After that that Data will be for various business purposes .To analyse the sales data for Big Manufacturer we can predict the sales using our modules .To analyse the Business Intelligence using power BI we can visualize sales prediction on various visualization techniques.

Robotic Process Automation has grown exponentially over the past few years. The demand for RPA is increasing in the RPA market as it promises to replace repetitive, rule-based, mundane, manual digital tasks with software robots. It also ensures organizations to make their operational processes error-free.

## REFERENCES

- [1]. Devansh Hiren Timbadia, Parin Jigishu Shah, Sugosh Sudhanvan, Supriya Agrawal. " Robotic Process Automation Through Advance Process Analysis Model" (IEEE 2020)
- [2]. Stefan Z., JovanoviÄ and S. Ä. Jelena. "Robotic Process Automation: Overview And Opportunities." International Journal" Advanced Quality"
- [3]. Ranjitha P, Spandana M "Predictive Analysis for Big Mart Sales Using Machine Learning Algorithms" (IEEE 2021)
- [4]. Cheng Zhang, junha Wen, and , Xiang gao, "An Improved Random Forest Algorithm for Predicting Employee Turnover", Hindawi Mathematical Problems in Engineering Volume 2019, Article ID 4140707
- [5]. Md. Anisur Rahman Mia\*, Mohammad Abu Yousuf , "Business Forecasting System using Machine Learning Approach" (IEEE 2021)
- [6]. Yada, Katsutoshi, Yuta, Kaneko, "A Deep Learning Approach for the Prediction of Retail Store Sales" , (IEEE 2016)
- [7]. Mohammed Ameer, Simhadri Prem Rahul, Dr.Suneetha Manne V.R.Siddhartha Engineering College Kanuru, Vijayawada-7 "Human Resource Analytics using Power Bi Visualization Tool " (IEEE 2020)
- [8]. Anand Singh Rajawat1 , Romil Rawat1 , Kanishk Barhanpurkar2 , Rabindra Nath Shaw3 and Ankush Ghosh4, " Robotic process automation with increasing productivity and improving product quality using artificial intelligence and machine learning ", ResearchGate January 2021
- [9]. Isabel Pedrosa, Jorge Bernardino, Bruno Varela, " Twitter Sensitivity Analysis in a Higher School Using Power BI " 2020 15th Iberian Conference on Information Systems and Technologies (CISTI) 24 – 27 June 2020, Seville, Spain ISBN: 978-989-54659-0-3
- [10]. M. L. Saini, Shilpi Kulshrestha, " Study for the Prediction of E-Commerce Business Market Growth Using Machine Learning Algorithm " 5 th IEEE International Conference on Recent Advances and Innovations in Engineering- ICRAIE 2020 (IEEE Record#51050)

# INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN SCIENCE, COMMUNICATION AND TECHNOLOGY

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal



## CERTIFICATE OF PUBLICATION

INTERNATIONAL STANDARD  
SERIAL NUMBER  
ISSN NO: 2581-9429

THIS IS TO CERTIFY THAT

**Ayush Mishra**

**Sinhgad Institute of Technology, Lonavala, Maharashtra, India**

HAS PUBLISHED A RESEARCH PAPER ENTITLED

**Robotic Process Automation with Increasing Productivity using Machine Learning  
IN IJAR SCT, VOLUME 3, ISSUE 2, MAY 2023**

Certificate No: 052023-A340  
[www.ijarsct.co.in](http://www.ijarsct.co.in)



# INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN SCIENCE, COMMUNICATION AND TECHNOLOGY

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal



## CERTIFICATE OF PUBLICATION

INTERNATIONAL STANDARD  
SERIAL NUMBER  
ISSN NO: 2581-9429

THIS IS TO CERTIFY THAT

**Rutik Kulkarni**

**Sinhgad Institute of Technology, Lonavala, Maharashtra, India**

HAS PUBLISHED A RESEARCH PAPER ENTITLED

**Robotic Process Automation with Increasing Productivity using Machine Learning  
IN IJAR SCT, VOLUME 3, ISSUE 2, MAY 2023**

Certificate No: 052023-A342  
[www.ijarsct.co.in](http://www.ijarsct.co.in)





# INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN SCIENCE, COMMUNICATION AND TECHNOLOGY

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal



## CERTIFICATE OF PUBLICATION

INTERNATIONAL STANDARD  
SERIAL NUMBER  
ISSN NO: 2581-9429

THIS IS TO CERTIFY THAT

**Govind Bhujbal**

**Sinhgad Institute of Technology, Lonavala, Maharashtra, India**

HAS PUBLISHED A RESEARCH PAPER ENTITLED

**Robotic Process Automation with Increasing Productivity using Machine Learning  
IN IJARSCT, VOLUME 3, ISSUE 2, MAY 2023**

Certificate No: 052023-A343  
[www.ijarsct.co.in](http://www.ijarsct.co.in)



# INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN SCIENCE, COMMUNICATION AND TECHNOLOGY

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal



## CERTIFICATE OF PUBLICATION

INTERNATIONAL STANDARD  
SERIAL NUMBER  
ISSN NO: 2581-9429

THIS IS TO CERTIFY THAT

**Sauraw Diwase**

**Sinhgad Institute of Technology, Lonavala, Maharashtra, India**

HAS PUBLISHED A RESEARCH PAPER ENTITLED

**Robotic Process Automation with Increasing Productivity using Machine Learning  
IN IJARSCT, VOLUME 3, ISSUE 2, MAY 2023**

Certificate No: 052023-A341  
[www.ijarsct.co.in](http://www.ijarsct.co.in)

