



---

## RESEARCH ON PREDICTIVE MODEL BASED ON CLASSIFICATION WITH PARAMETERS OF OPTIMIZATION

*T. Gulzat\**, *N. Lyazat\**<sup>†</sup>, *V. Siladi*<sup>‡</sup>, *S. Gulbakyt*<sup>§</sup>, *S. Maxatbek*<sup>‡</sup><sup>¶</sup>

---

**Abstract:** This paper effectively uses the data mining and optimization methods to investigate a classification based on decision trees algorithm, then optimizes by the method of grid search and cross-validation, which improves the prediction accuracy of the decision tree model for the PCs sales in practical application and solves insufficient training data, high computational cost, and low prediction accuracy. The main goal of the article is to predict PC sales using machine learning tools caused by various types of operating system factors in practical applications. This article proposes a combined predictive research model that fully reveals the benefits of optimization and neural networks, and also has a very accurate fit and forecasting accuracy. The proposed predictive model is implemented in the data science software platform RapidMiner. A decision tree model is executed, then the model's prediction capacity is evaluated and tested. Grid search optimizer is used to automatically build the final model using the best-optimized parameter for training the classifier. The paper combines grid the grid search and cross-validation to optimize the parameters of the decision tree to improve the classification prediction accuracy of the decision tree model. This article combines neural networks with optimization methods to establish a prediction model for laptop sales. This model gives full play to the advantages of optimization and neural networks and has very good fitting capabilities and prediction accuracy. Besides, the neural network for the prediction model has strong dynamic analysis capabilities. Once there are new observations, it can continue to be added to the modeling, which has high adaptability. The Neural Network algorithm has the highest accuracy of the predicted PC sales by evaluating the results of the five kinds of algorithms. The result for prediction accuracy shows the highest performance.

Key words: *Machine Learning (ML)*, *optimization*, *predictive modeling*, *Data Mining (DM)*, *decision tree*

---

\*Turken Gulzat – Corresponding author; Naizabayeva Lyazat; Al-Farabi Kazakh National University, Almaty, Kazakhstan, E-mail: [turken.gulzat@gmail.com](mailto:turken.gulzat@gmail.com)

<sup>†</sup>Naizabayeva Lyazat; Vladimir Siladi; Sembina Gulbakyt; Al-Farabi Kazakh National University, International Information Technology University, Almaty, Kazakhstan, E-mail: [naizabayeva@gmail.com](mailto:naizabayeva@gmail.com)

<sup>‡</sup>Vladimir Siladi; Satymbekov Maxatbek; Matej Bel University, Banska Bystrica, Slovakia, E-mail: [vladimir.siladi@umb.sk](mailto:vladimir.siladi@umb.sk)

<sup>§</sup>Sembina Gulbakyt; International Information Technology University, Almaty, Kazakhstan, E-mail: [gsembina01@gmail.com](mailto:gsembina01@gmail.com)

<sup>¶</sup>Satymbekov Maxatbek; Al-Farabi Kazakh National University, Almaty, Kazakhstan, E-mail: [m.n.satymbekov@gmail.com](mailto:m.n.satymbekov@gmail.com)

*Received: May 6, 2019*

DOI: 10.14311/NNW.2020.30.020

*Revised and accepted: October 30, 2020*

## 1. Introduction

In data mining methods, optimization techniques play an extremely important role as the level of achievement increases. It further enhances computational efficiency. Therefore, the problem of increasing the prediction accuracy about the product is essential to the success of the industry in satisfying the consumer demand and increasing the competitiveness of enterprises as different factors will cause a different rate of PC sales.

Nowadays, with the fast improvement of the commodity economy, the increasing diversification of customers and the demand for individualization have led to a significant increase in the difficulty of controlling the production and sales of goods. Effective optimization methods play a vital role in sales forecasting. Their effects are significant in improving the overall performance of the whole organization. Vikas Kumar writes about this in his article “JIT Based Quality Management: Concepts and Implications in Indian Context” [1]. When the enterprise is planning for increasing its productivity during financial investment, frequently encountered problems are how to use the limited resources to maximize the profit margin and decrease the loss. Likewise, there is an optimization problem in the application process: how to optimize resources to enhance the enterprise’s profits in determining types of products. In these cases, the solution of the prediction accuracy problem, which involves the product type and quantity to produce, also considers customer behavior analysis.

Predictive modelling is described as the methodological framework for defining and solving systematic value investing scenarios as optimization problems in the work of R.J. Sak: An Optimized and Dynamic Solution Framework for Systematic Value Investing [2]. Predictive models allow us to take what we already know about the behavior of a system and use it to predict how that system will behave in new circumstances, Scott P. Stevens writes about this in Mathematical Decision Making: Predictive Models and Optimization [3]. Various data mining tools and machine learning algorithms are supplied in the predictive analysis model, some of these algorithms were developed to solve specific business problems, enhance existing algorithms, or provide new capabilities which may make some of them more appropriate for purposes.

Comparing different runs of different algorithms always brings surprising findings. By doing this, it gives more detailed insight into the business and supports identify which variables within data have predictive power. A predictive model can take various algorithms like Decision tree, SVM, Naive Bayes Classification, K-NN, Random Forest, Neural Network, etc. Besides, certain models work best for certain data and analyses, also using classification algorithms that employ decision rules decide the outcome of a given scenario or transaction.

The potential applications of data mining techniques are very broad; it can be utilized in all areas of business operations, scientific research, and industrial enterprise management, and government decision-making. Nevertheless, the final

goal of data mining is in the optimization of business decisions and behaviors based on the analysis results obtained by data mining tools, thereby increasing the efficiency of enterprises. Unfortunately, in practical applications, traditional, general data mining technology only provides a set of data. The interpretations of mining results, deployment, etc. are rarely involved, so there is a lack of certain business decisions and behavior optimization and classification capabilities. This limits a large number of applications for having clear business goals. For these reasons, the combination of optimization methods and data mining techniques for the predictive model is most effective; they complement and use each other, which provide an effective solution.

The application effect of the optimization method in data mining is reflected in the following aspects:

- 1) The core data mining is directly defined as an optimization problem, which is solved by the optimization method to improve the accuracy of the algorithm. Support Vector Machine is a typical example of the optimization method to solve the classification problem. Directly defining a data mining problem as an optimization problem will often provide new threads and solutions.
- 2) In the complex data mining problem, the optimization method is used to solve one of the sub-problems to achieve global optimization. Typical examples include preprocessing of data and post-processing of rules, just like complex optimization problems are described in the article [4].
- 3) Directly defining complex data mining problems as a complex optimization problem, and obtaining approximate solutions by using data such as genetic algorithms. In this regard, the combination of "customer segmentation" and "marketing strategy deployment" as a complex optimization problem is a good example, such as the article discussed [5].
- 4) Selecting the data mining model by optimization method. For the same data mining problem, multiple different models can be built at the same time, and the better ones are selected by the optimization technique as the most middle model.

In this study, based on the above the fourth aspect seeks to develop optimization that can support the predicting model. Hence, the product predicting model is developed by using machine learning methods to get the approach. The purpose of this article is to build a predictive model of PC sales, namely, to obtain a process with multiple decisions about response variables, machine learning algorithms, predictor variables, and selection of criteria. We will use supervised machine learning tools to predict PC sales caused by different types of operating system factors in practical applications.

Building a predictive model is a process with several decisions about response variables, machine learning algorithms, predictor variables, and selection criteria.

The proposed forecasting model will be implemented in the Rapid Miner software platform for data analysis. Therefore, to achieve the goal, it is necessary to solve the following tasks:

- Preparing a dataset and select a data mining tool,
- Processing raw data, pre-process the dataset with normalization,
- Classifying machine learning methods (cross-validate, do parameter optimization),
- Making a forecast of the model,
- Drawing a model conclusion and get the analysis results.

## 2. Related work

Several scientific researchers have worked with advanced methodologies for creating optimization and prediction by using machine learning algorithm, the results of studies are briefly discussed below. Md Tahsir Ahmed Munna, Mirza Moh-tashim Alam used the Artificial neural network (ANN) and K-mean clustering methods, for improving statistical prediction accuracy of chronic diabetes mellitus based on extensive classification system, allow patients to know in advance about the prevalence of the disease [6]. Youshang Zhang, Qi Li utilized the regressive convulsive neural network (RCNN) and the reference support vector regressive method (SVR), considering the account the factors affecting electricity consumption, created a prediction model with a low error rate [7]. In the next job authors identified many issues influenced the achievements of university students and compared different classification algorithms, resulting in a high-performance classifier Naive Bayes, which offered high-performance calculation to help optimize of resource utilization [8]. In another article, they tested mobile phone metadata with different algorithms like K-nearest neighbors (K-NN), random forest, Support Vector Machine (SVM), developed a framework which allowed data to be augmented to predict useful individual characteristics and group estimates [9]. This model was developed using big data and machine learning algorithm to determine the gender and age of the user based on mobile phone data, in order to increase the accuracy of the solution corresponding to consumer data in telecommunication and showed a high performance in the prediction process [10]. The analysis used several machine learning algorithms, performance for the statistical analysis of traffic predict model represented minimum accuracy [11]. The ordinal prediction model studied the problem of final nonlinear optimization based on the principles of machine learning and optimization theory, the research showed a higher result in the form of ranking [12]. Stephen Gang Wu and other researchers utilized three machine learning methods, studied the relations between influencing factor and metabolic flow. Achieved a grid search for the best parameter of each algorithm and verified its performance to predict bacterial central metabolism [13]. In [14] and [15] described quantum entanglement as the main feature of wave probabilistic model in data analysis. Based on the above research, we can see the remarkable achievements of optimization and prediction through machine learning algorithm in various fields around the world.

### 3. Brief review of methodologies

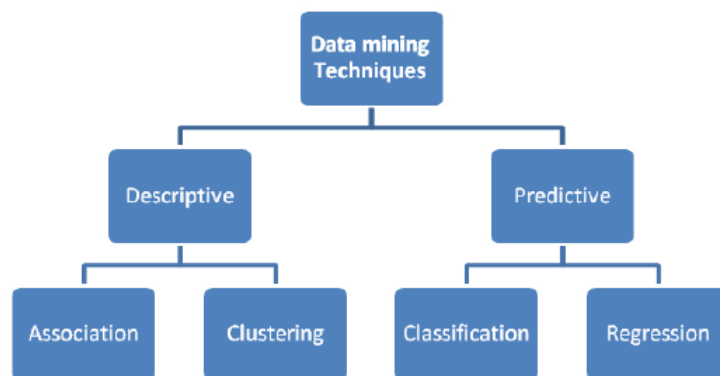
Rapid developments in information technology have resulted in the construction of many business intelligence (BI) systems in various areas. Data have become a pivotal resource in many organizations, consequently, efficacious accessing, sharing the data, and extracting important information from the data, as well as making use of the stored information has turned into an urgent need. As a result, there have been many enterprises on firstly incorporate various data sources such as databases to build a corporate data warehouse. Predicting information from the data warehouse in the form of patterns and trends by applying the data mining tools is more apparent.

Statistical methods only involve determining explicit information; they do not involve any validation mechanism. However, data mining methods can discover implicit knowledge through data analysis.

We decided to use supervised machine learning tools to predict PC sales caused by different types of operating system factors in practical applications.

#### 3.1 Data Mining technologies

With the continuous collection and storage of a large amount of data, many industries are beginning to be interested in mining to extract significant patterns from their databases. Finding interesting relationships in a large number of business transaction records can help with many business decision-making and predicting processes, such as catalog design, cross-marketing, and customer shopping behavior analysis. Data mining is a multidisciplinary field, drawing work from areas including database technology, machine learning, statistics, pattern recognition, information retrieval, neural networks, knowledge-based systems, artificial intelligence, high-performance computing, and data visualization [16]. A large number of good techniques for mining operations are shown in Fig. 1.



**Fig. 1** *Data Mining Techniques.*

The Fig. 1 shows the scheme proposed by the authors in the work: Popular Decision Tree Algorithms of Data Mining Techniques [17].

All kinds of organizations can construct machine learning models to address higher productivity than ever before on the platform, the Data Mining software tools integrate basic principle, algorithms, concept, and approaches, furthermore data mining applications are based on their operation in an algorithm to find patterns of data for investigation and analysis of data using a set of tools that permit the visualization of results and reports. Five modern, open-source data mining tools are presented and compared in [18], such as WEKA, Microsoft Azure ML Studio, RapidMiner, H2O, R and Apache Spark, etc.

RapidMiner is a combination of priority open-source data mining tools with software Java. RapidMiner eliminates the complexities of cutting edge data science by making it easy to deploy the latest machine learning algorithms and innovative technologies like Tensorflow, Hadoop, and Spark. More than 300,000 data scientists in over 150 countries use RapidMiner products on premise or in the cloud to drive revenue, reduce costs, and avoid risks [19]. In this study, we utilize RapidMiner studio, because it has enormous flexibility in process design (e.g. considering the ability of nesting optimization, evaluation, etc.). RapidMiner is deployed for different predictive analytics use-cases, such as churn prevention, predictive maintenance, risk modeling, text mining, and customer segmentation.

### 3.2 Machine Learning

Machine Learning (ML) and intelligent systems have become an indispensable part of our modern society [20]. Most of these systems are according to recognizing patterns in observable data, to understand the data or make new predictions on unseen data. Machine Learning Technology utilizes experimental data to optimize clustering, classification of parameters to support and enhance excellently models that proposing for predict behavior or properties of systems. Nevertheless, machine learning techniques such as Regression and Classification Trees, Random Forests, Gradient Boosted Machines, and Deep Learners have recently acquired increasing popularity in many business applications.

**Decision tree** Trees are used in data structures, databases, computer algorithms, machine learning, and data mining. This method produces a non-parametric classification and prediction model [21]. It is a reliable and effective algorithm that provides high classification accuracy with a simple representation of gathered knowledge. A predictive modeling technique is utilized in the machine learning process such as classification, clustering, and predictive process to find the optimal decision tree by minimizing the generalization error.

**Naïve Bayes** Naive Bayes classifier (NB, Hand and Yu, 2001) technique, based on the so-called Bayesian theorem classification, is named after Thomas Bayes (1702–1761) and is particularly suited when the Trees dimensionality of the inputs is high. Despite its simplicity, Naive Bayes can often outperform more sophisticated classification methods.

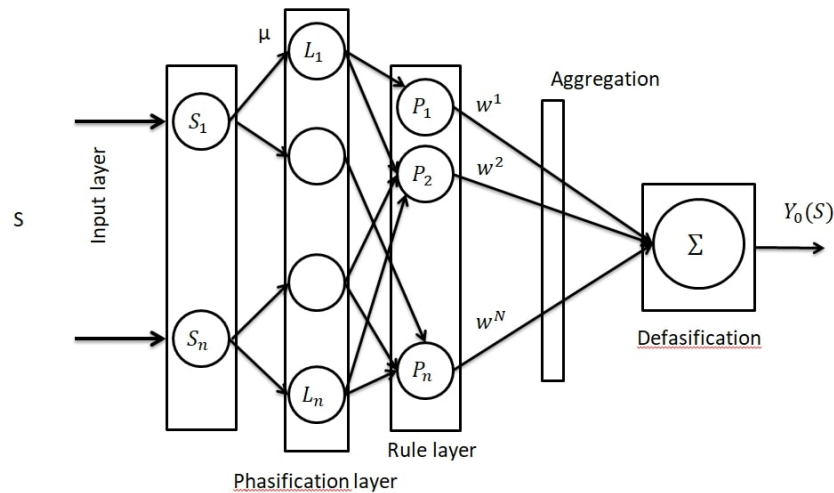
**K-nearest neighbors** K-nearest neighbors (K-NN) algorithm is a simple, easy to implement supervised machine learning algorithm that can be used to predict the

same value/class as the nearest instance in the dataset and can solve classification problems.

**Random Forest** Random Forest has been widely used in classification and prediction. Random Forests are a combination machine learning algorithm which is combined with a series of tree classifiers. Each tree cast a unit vote for the most popular class, then combining these results gets the final sort result [21].

**Neural Network** Currently, there are many machine learning algorithms for creating a scoring model, such as the decision tree, K-NN (K-nearest neighbor method), SVM (support vector machine), and NN (neural network) [22].

How clear should the decisions that affected the model's results be? In other words, how important is it for us to explain the structure of the model? We decided to choose a neural network in choosing a machine learning algorithm. At this time, there are many great structures, such as TensorFlow, Theano. Fig. 2 shows how these structures make it possible to make deep and effective use of their architectural and training parameters. Neural networks are an effective method for finding errors in nonlinear networks.



**Fig. 2** Neural network layers.  $w$  – sum of weight,  $Y_0(S)$  – control signal,  $\mu$  – membership function.

$$S = (S_1, \dots, S_n), \tag{1}$$

a vector describing the current state of incoming variables is passed to the network input  $L$ .  $S_j$  linguistic values for the input variable  $i$ .  $P$  – neural network function generator.  $\Sigma$  – neural summation.

Due to the presence of artificial associative layers, an artificial neural network can build hypotheses based on the search for complex dependencies.

### 3.3 Research frame work

The predictive modeling framework proposed in this research shown in Fig. 3.

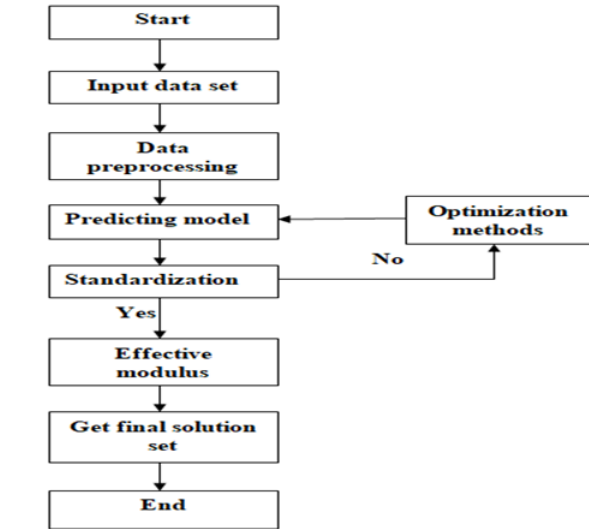


Fig. 3 Flow chart of building a predictive model.

## 4. Analysis and experiments

Building a predictive model is a process with several decisions about response variables, machine learning algorithms, predictor variables, and selection criteria. Due to the combinational possibilities of various model specifications, it is not efficient to evaluate an exhaustive list of every possible model specification.

When creating the predictive model, the RapidMiner technology procedure provides the following process:

- 1) Preparation of data set and selection of a data mining tool,
- 2) Raw data processing,
- 3) Preprocessing the data set with normalization,
- 4) Classification with Machine learning methods (Cross-validation, parameters optimization),
- 5) Prediction of the model,
- 6) The output from the model and results of the analysis.

The proposed predictive model is implemented in RapidMiner (data science software platform) [23] to achieve the goal and effective results. As a source of



information to perform the analysis, we use a dataset “Laptop Price” [24]. The dataset includes 12 regular and 1 special attributes, 1,303 variables. These are Company, Product, TypeName, Inches, ScreenResolution, Cpu, Ram, Memory, GPU, OpSys, Weight, Price euros.

The predictive model requires a dataset with complete observations. Missing data is a problem affecting most databases and electronic medical records (EMR) are no exception. Because most statistical models operate only on complete observations of exposure and outcome variables, it is necessary to deal with missing data, either by deleting incomplete observations or by replacing any missing values with an estimated value based on the other information available [25].

After the data set is loaded into the RapidMiner repository, in preprocessing steps, the latter parameter is applied for completing missing categorical and numeric attributes with a given attribute.

A ROC (Receiver Operating Characteristic) graph is a technique for visualizing, organizing, and selecting classifiers based on their performance. Algorithms ROC graphs have long been used in signal detection theory to depict the trade between hit rates and false alarm rates of classifiers. One of the earliest adopters of ROC graphs in machine learning was Spackman (1989), who demonstrated the value of ROC curves in evaluating and comparing algorithms [26]. Thus, it is used as a post-processing technique and as a method that is actively taking part in the model construction to improve a given model. Fig. 4 below is a view of the

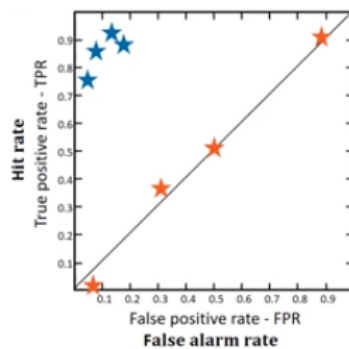
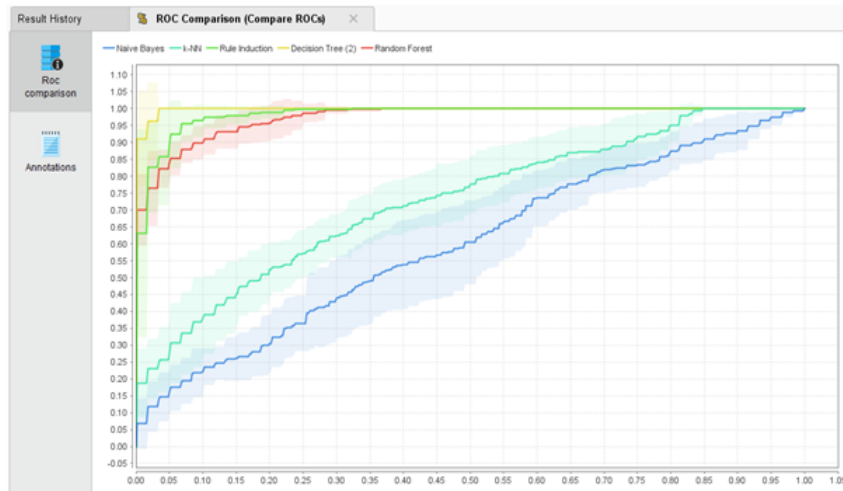


Fig. 4 A basic ROCs graph.

basic ROC curve concept. MOD applicable model is using the Compare ROCs operator (<http://mod.rapidminer.com>) based on ROC graph concept to analysis and compares the most frequently used algorithms, such as Decision tree, Naive Bayes, K-NN, Rule Induction, Random Forest algorithms. Then choose the best algorithm for classification to optimize the parameters. The result has shown in the basic ROCs graph (Fig. 5).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}}, \quad (2)$$

where TP – true positive, FP – false positive, FN – false negative, TN – true negative counts.



**Fig. 5** Result of ROCs graph.

According to the result of the ROC Comparison, we can see that those decision tree algorithms behave with the highest accuracy. We perform a decision tree model, then apply, score, and test the prediction capacity of the model.

Optimization hyper parameter Grid-searching is the process of scanning the data to configure optimal parameters for a given model. Grid search optimizer is used to automatically build the final model using the best-optimized parameter for training the classifier. Cross-validation (CV) is a statistical method used to verify the performance of the classifier. We combine the grid search and cross-validation to optimize the parameters of the decision tree to improve the classification prediction accuracy of the decision tree model. After optimizing the parameters of the grid search, we can get higher accuracy in the test set.

In this paper, when using the cross-validation method, the learning data samples need to be divided into two parts: training data samples and verification data samples. The first part of the training set is to train the classifier, and then to use the verification set to test the training model as the evaluation classifier performance indicators. Subsequently, connecting of the input and output in the “cross-validation” operator, we define the parameters which we want to optimize. The results can be seen here Fig. 6).

## 5. Discussion and result

In this paper, we have configured 404 parameter combinations and 10 fold cross-validation in each loop. Based on these, we are producing over 4,000 decision trees. As shown in the following result, the minimal gain is close to the default of the decision tree which is set to 0.089. The criterion “accuracy” for the operating system – Windows 10 showed a higher accuracy. It also briefly created a chart displaying the parameters that we used and the performance of different combinations.

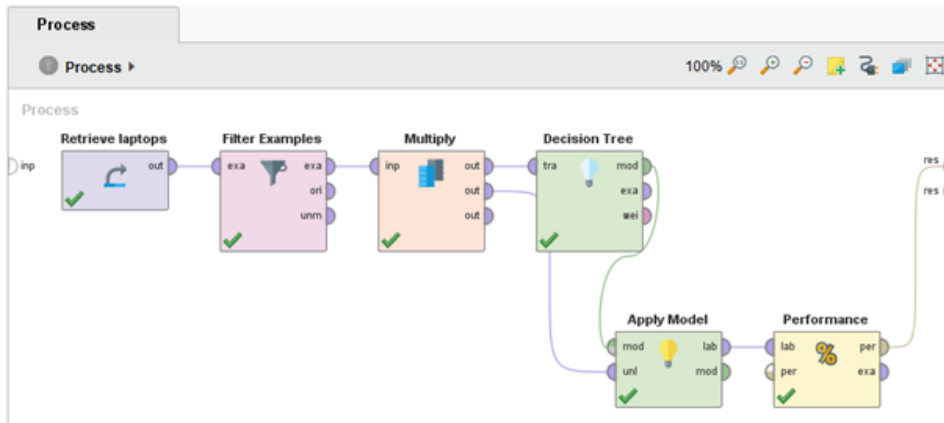


Fig. 6 Process of apply and test for Decision Tree Model.

Prediction performance reached 84% with optimized parameters (Fig. 7, 8).

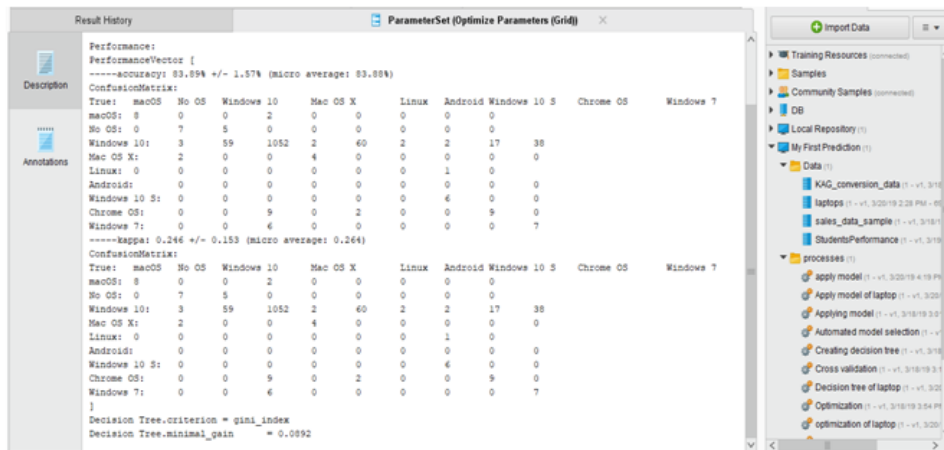


Fig. 7 Prediction performance.

Classification performance without focussing on a class is the most general way of comparing algorithms [27].

As shown in the Fig. 9, the accuracy metric was compared between models. As a result of the comparison, the accuracy metric of the neural network shows 88%, while its error is only 12%. If you look at other methods, the accuracy of the metric is lower than in the neural network, and their errors will be higher accordingly.

## 6. Conclusion and future work

Due to changes in demand and competition situation, the classification of each project, differentiation is bound to happen in a changing market environment. If



Fig. 8 Visualization of operating system.

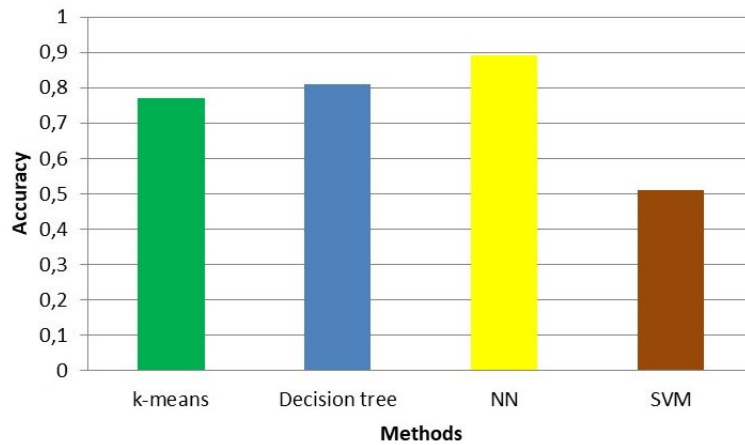


Fig. 9 Comparison of data analysis methods.

enterprises do not pay attention to the development of new products and the elimination of declining products, there will be an unsound and unbalanced product classification. To this end, companies need to constantly analyze the sales growth rate, profit margin, and market share of each product line and determine the potential development trend of each product item growth to determine the use of corporate funds. Prediction makes decisions to develop new products and eliminate recession products to adjust their product classification and optimization.

This article combines neural networks with optimization methods to establish a prediction model for laptop sales. This model gives full play to the advantages of

optimization and neural networks, and has very good fitting capabilities and prediction accuracy. Besides, the neural network for the prediction model has strong dynamic analysis capabilities. Once there are new observations, it can continue to be added to the modeling, which has high adaptability.

The Neural Network algorithm has the highest accuracy of the predicted PC sales by evaluating the results of the five kinds of algorithms. The result for prediction accuracy shows the highest performance with about 88%. The proposed model helped analyze PC sales caused by the different types of operating system factors. Combined with this model, we can further study how different variables in the operating system links affect the PC sales, but also can be more detailed sub-regional, divided into different classes for research.

In future work, we will analyze other publicly available datasets and more advanced architectures, and will improve the selection of the best parameters to further enhance the predictive accuracy. This will help to enable more effective production planning.

## References

- [1] KUMAR V. JIT Based Quality Management: Concepts and Implications in Indian Context. *International Journal of Engineering Science and Technology*. 2010, 2(1), pp. 40–50.
- [2] SAK R.J. Predictive Modeling: An Optimized and Dynamic Solution Framework for Systematic Value Investing. *Social Sciences Computing cluster (SSCC) at Northwestern University*. 2010, 2(1), pp. 3.
- [3] STEVENS S.P. Mathematical Decision Making: Predictive Models and Optimization. *The Great Courses*. 2015, 24(2), pp. 10–12.
- [4] SERBAN A. Fast and robust Matlab-based finite element model used in the layup optimization of composite laminates. *7th International Conference on Advanced Concepts in Mechanical Engineering*. 2016, 147, doi: [10.1088/1757-899X/147/1/012103](https://doi.org/10.1088/1757-899X/147/1/012103).
- [5] Ivett Fuentes, Gonzalo Napoles, Leticia Arco, Koen Vanhoof. Customer Segmentation using Multiple Instance Clustering and Purchasing Behaviors. *Research Gate*. 2018, doi: [10.1007/978-3-030-01132-1\\_22](https://doi.org/10.1007/978-3-030-01132-1_22).
- [6] MUNNA M.T.A., ALAM M.M., ALLAYEAR M., SARKER K., ARA S.J.F. Prediction Model for Prevalence of Type-2 Diabetes Complications with ANN Approach Combining with K-Fold Cross Validation and K-Means Clustering. *FICC 2019: Advances in Information and Communication*. 2019, pp. 1031–1045, doi: [10.1007/978-3-030-12388-8\\_71](https://doi.org/10.1007/978-3-030-12388-8_71).
- [7] ZHANG Y., LI Q. A Regressive Convolution Neural Network and Support Vector Regression Model for Electricity Consumption Forecasting. *FICC 2019: Advances in Information and Communication*. 2019, pp. 33–45, doi: [10.1007/978-3-030-12385-7\\_4](https://doi.org/10.1007/978-3-030-12385-7_4).
- [8] MUCHUCHUTI S., NARASIMHAN L., SIDUME F. Classification Model for Student Performance Amelioration. *FICC 2019: Advances in Information and Communication*. 2019, pp. 742–755, doi: [10.1007/978-3-030-12388-8\\_51](https://doi.org/10.1007/978-3-030-12388-8_51).
- [9] JAHANI E., BJELLAND J., BENGTSSON L., DE MONTJOYE Y.A. Improving ofcial statistics in emerging markets using machine learning and mobile phone data. *EPJ Data Science*. 2017, 6(3), pp. 1–21, doi: [10.1140/epjds/s13688-017-0099-3](https://doi.org/10.1140/epjds/s13688-017-0099-3).
- [10] MOUSA I., JAFAR A., ALJOUAA K. Predicting customer's gender and age depending on mobile phone data. *Al Zuabi et al. J Big Data*. 2019, 6(18), pp. 1–16, doi: [10.1186/s40537-019-0180-9](https://doi.org/10.1186/s40537-019-0180-9).
- [11] ADEKITAN A.I., ABOLADE J., SHOBAYO O. Data mining approach for predicting the daily Internet data traf of a smart university. *Adekitan et al. J Big Data*. 2019, 6(11), pp. 2–24, doi: [10.1186/s40537-019-0176-5](https://doi.org/10.1186/s40537-019-0176-5).

- [12] SADER M., VERWAEREN J., PEREZ-FERNANDEZ R., DE BAETS B. Integrating expert and novice evaluations for augmenting ordinal regression models. *ScienceDirect*. 2019, 51, pp. 1–9, doi: [10.1016/j.inffus.2018.10.012](https://doi.org/10.1016/j.inffus.2018.10.012).
- [13] WU S.G., WANG Y., JIANG W., OYETUNDE T., YAO R., ZHANG X., SHIMIZU K., TANG Y.J., BAO F.S. Rapid Prediction of Bacterial Heterotrophic Fluxomics Using Machine Learning and Constraint Programming. *PLOS Computational Biology*. 2016,12(4), pp. 1–22, doi: [10.1371/journal.pcbi.1004838](https://doi.org/10.1371/journal.pcbi.1004838).
- [14] SVITEK M. Theory and Algorithm for Time Series Compression. *Neural Network World*. 2005, 15(1), pp. 53–67.
- [15] SVITEK M. Complementary variables and its application in statistics. *Neural Network World*. 2007, 17(3), pp. 237–253.
- [16] KUMAR A., TYAGI A.K., TYAGI S.K. Data Mining: Various Issues and Challenges for Future A Short discussion on Data Mining issues for future work. *International Journal of Emerging Technology and Advanced Engineering*. 2014, 4(1), pp. 1.
- [17] RADHWAN H.A., ALSAGHEER, ALHARAN A.F.H., AL-HABOABI A.S.A. Popular Decision Tree Algorithms of Data Mining Techniques: A Review. *International Journal of Computer Science and Mobile Computing*. 2017, 6(6), pp. 133–142.
- [18] DUSANKA D., DARKO S., SRDJAN S., MARKO A., TEODORA L. A Comparison of Contemporary Data Mining Tools. *XVII International Scientific Conference on Industrial Systems*. 2017, pp. 1–7.
- [19] RAPIDMINER FOUNDATION, RapidMiner Studio 9.6.000 [software]. 2006 [accessed 2019-02-14]. Available from: <https://rapidminer.com>
- [20] REDDI S.J. New Optimization Methods for Modern Machine Learning. *School of Computer Science Carnegie Mellon University Pittsburgh*. 2016, pp. 5.
- [21] CINAROGLU S. Comparison of Performance of Decision Tree Algorithms and Random Forest: An Application on OECD Countries Health Expenditures. *International Journal of Computer Applications*. 2016, 138(1), pp. 37–41, doi: [10.5120/ijca2016908704](https://doi.org/10.5120/ijca2016908704).
- [22] LIU Y., WANG Y., ZHANG J. New Machine Learning Algorithm: Random Forest. *ICICA: Information Computing and Applications*. 2012, 7473, pp. 246–252.
- [23] TEREIKOVSKIY I., TEREIKOVSKA L., KORYSTIN O., MUSSIRALIYEVA S., SAMBETBAYEVA. A Keystroke Authentication and Recognition of Emotions Based on Convolutional Neural Network. In: *Advances in Artificial Systems for Medicine and Education*. AIMEE 2019: pp. 283–292, III. doi: [10.1007/978-3-030-39162-1\\_26](https://doi.org/10.1007/978-3-030-39162-1_26).
- [24] RAPIDMINER [Data science platform] Available from: <https://rapidminer.com/get-started/>.
- [25] KAGGLE [A data modeling and data analysis competition platform]. 2010-04 [accessed 2018-04-03]. Available from: <https://www.kaggle.com/datasetssortBy=relevance&group=public&search=Laptop+Price+&page=1&pageSize=20&size=all&filetype=all&license=all>.
- [26] SALGADO C., VIEIRA S. Missing Data. *Secondary Analysis of Electronic Health Records*. 2016, pp. 143–162, doi: [10.1007/978-3-319-43742-2\\_13](https://doi.org/10.1007/978-3-319-43742-2_13).
- [27] FAWCETT T. ROC Graphs: Notes and Practical Considerations for Researchers. *ResearchGate*. 2004, 31(8), pp. 1–38.
- [28] SOKOLOVA M., JAPKOWICZ N., SZPAKOWICZ S. Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation, *ResearchGate*, 2014, doi: [10.1007/11941439\\_114](https://doi.org/10.1007/11941439_114).