**Analysis of Textile stock prices Authentication System Using Different Machine Learning Techniques**

# SIDDHARTH DOGRA RISHU VERMA

# Business Analytics and Big Data Business Analytics and Big Data

# Thapar University Thapar University

# 

# 

***ABSTRACT:***This study focuses on investigating the real-world uses of data mining in the textile stock pricesindustry, with a particular emphasis on classification, clustering, and machine learning approaches. Notably, support vector machines and artificial neural networks are widely used and exhibit good accuracy rates in textile stock pricesapplications. The K-means technique is frequently used in clustering problems, among other options. The essay highlights the usefulness of data mining methods in the textile stock pricessector and suggests future research directions. In addition, a novel framework for multi-objective optimization of textile stock pricesmanufacturing processes is introduced. This framework uses the deep Q-networks algorithm in a multi-agent system to solve the optimization problem by turning it into a stochastic game. The suggested methodology outperforms conventional approaches and successfully addresses the difficulties brought on by the big data era in textiles. Lastly, a multilayer, multi-attentional deep learning network is suggested to handle the challenging task of identifying multilabel faults in textile stock pricesphotos. When compared to existing methods, this network performs exceptionally well in extracting different features and precisely identifying a variety of faults.

*.*

**KEYWORDS:** Textile stock prices Authentication, Logistic Regression, Naïve Bayes, Decision Tree, Support Vector Machine (SVM), Random Forest, K-Nearest Neighbor (KNN)

# 1.0 INTRODUCTION

In today's highly competitive business environment, accurate forecasting of a company's performance plays a crucial role in decision-making and strategic planning. In the textile stock pricesindustry, where market dynamics are influenced by various factors such as changing consumer preferences, fluctuating raw material costs, and evolving fashion trends, effective performance forecasting becomes even more critical. Traditional forecasting methods often fall short in capturing the complex patterns and nonlinear relationships present in textile stock pricescompany data. However, with the advancements in machine learning, specifically regression techniques, there is an opportunity to improve the accuracy and reliability of performance forecasts.

This research paper aims to explore the application of machine learning regression methods for forecasting the performance of textile stock pricescompanies. By leveraging historical data, we can train regression models that capture the underlying patterns and relationships between key performance indicators (KPIs) and various market factors. These models can then be used to generate accurate forecasts, enabling textile stock pricescompanies to make informed decisions, optimize resource allocation, and improve overall business performance.

# 2.0 LITERATURE SURVEY

**Forecasting Textile Company Performance using Machine Learning Regression Methods**

Preserving the authenticity of higher denomination printed textile stock prices is one of the critical issues. It has a major role in the financial activities of every country. A variety of models are proposed for textile stock prices authentication through different approaches and using different machine-learning techniques.

The study proposed by Kumar *et al*.,2005 evaluates different machine learning techniques such as Probabilistic Neural Network (PNN), Multi-layer Perceptron (MLP), Radial Basis Function (RBF), Decision Tree (DT), Naïve Bayes and concludes that Decision-Tree and MLP technique is best to classify a textile stock prices.

A study proposed by Shahani *et al*., 2018 uses Back Propagation Neural Network (BPN) and Support Vector Machine and concludes that Back Propagation Neural Network gives 100% accuracy in detecting forged currency notes. In the proposed study by Nastoulis *et al.,* 2006 a new method is proposed for textile stock prices recognition; Probabilistic Neural Network (PNN). The study by Omatu *et al.,* 2007 uses LVQ classifier for textile stock prices recognition. The experiment has been applied to US dollars and can be used for other kinds of textile stock pricess.

In Gillich *et al*., 2014 the features of the textile stock prices are extracted using Fast Wavelet Transforms. Later, one-against-all classification approach was employed that classifies the note into four different categories: Genuine, High-Quality Forgery, Low-Quality Forgery, and Inappropriate ROI which resulted in 100% detection rate.

In Rana *et al.,*2021 uses pattern recognition and image processing learning and analyzing methods for

distinguishing features.

The study proposed by Lalita *et al.,* 2014 uses recognition system consists of preprocesses,

the NN system, and the DSP unit. On the slab values extraction, the slab values, which represented as characteristics of textile stock prices, are extracted from each textile stock prices image by using an axis-symmetry mask set.

In Aoba *et al.,* 2003 propose a textile stock prices recognition system composed of two parts; a classification part and a validation part. The classification part uses a three-layered perceptron and the validation part uses several RBF networks.

The study proposed by Mohamad *et al.,* 2014 aims to classify the sample of textile stock pricess data into the MATLAB’s GUI application which is to examine

whether the data is real or forged. The tools that will be used in this research is ‘nntool’ that stands for Neural Network technique. Neural network is used to simulate research, develop and apply artificial neural networks, biological neural networks and in some cases a wider array of adaptive systems.

In Khairy *et al.,* 2020 detection of textile stock pricess is proposed based on ensemble techniques of Boosting and Voting algorithms used with a selection of ten (10) standard machine learning algorithms. It adopted ensemble learning algorithms of AdaBoost and Voting (with a combination rule of the average of probabilities). The integration of the machine learning algorithms into the ensemble methods result in a classification outcome for effective detection.

Table 1: Summary of related work

|  |  |  |  |
| --- | --- | --- | --- |
| References | Technique Used | Dataset Used | Performance Measures |
| Kumar *et al*.,2005 | * Probabilistic neural network. * Multi-layer Perceptron * Radial Basis Function * Decision Tree * Naïve Base | [https://archive.ics.uci.edu/ml/machine-learning-databases/00267/data\_textile stock prices\_authentication.txt](https://archive.ics.uci.edu/ml/machine-learning-databases/00267/data_banknote_authentication.txt) | Sensitivity  Specificity  Accuracy |
| Shahani *et al*., 2018 | * Back Propagation Neural Network * Support Vector Machine | [https://archive.ics.uci.edu/ml/machine-learning-databases/00267/data\_textile stock prices\_authentication.txt](https://archive.ics.uci.edu/ml/machine-learning-databases/00267/data_banknote_authentication.txt) | Accuracy  Sensitivity  Specificity  Precision |
| Nastoulis *et al.,* 2006 | * Probabilistic Neural Network (PNN) | Image pre-processing | - |
| Omatu *et al.,* 2007 | Learning vector quantization | Image pre-processing | Reliability |
| Gillich *et al.,* 2014 | * Wavelet Transform * Intaglio | Heterogeneous textures  Extraction by using image pre-processing | RMSE (root mean squared error) |
| Rana *et al.,* 2021 | * Support Vector Classifier * Gradient Boosting Classifier | [https://archive.ics.uci.edu/ml/machine-learning-databases/00267/data\_textile stock prices\_authentication.txt](https://archive.ics.uci.edu/ml/machine-learning-databases/00267/data_banknote_authentication.txt) | Accuracy  F1-Score |
| Lalita *et al.,* 2014 | * Neural Network | Image pre-processing | - |
| Aoba *et al.,* 2003 | * Three- layered perceptron * Radial Basis Function (RBF) networks | Image pre-processing | - |
| Mohamad *et al.,* 2014 | * Artificial Neural Network (ANN) | [https://archive.ics.uci.edu/ml/machine-learning-databases/00267/data\_textile stock prices\_authentication.txt](https://archive.ics.uci.edu/ml/machine-learning-databases/00267/data_banknote_authentication.txt) | - |
| Khairy *et al.,* 2020 | * AdaBoost * Voting | - | Accuracy  detection rate  Matthews  correlation coefficient |

# 3.0 DATASET DESCRIPTION

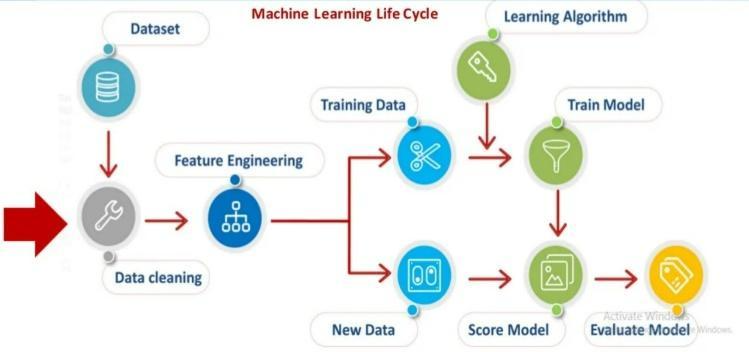
The dataset used to train the models is taken from UCI machine learning repository. Data was extracted from genuine and counterfeit textile stock prices images. The dataset has 1372 instances. There are 5 attributes out of which 4 are the features and one is the target attribute. The dataset contains a balanced ratio of both classes which is 55:45(genuine: counterfeit). The target class contains two values: 0 and 1 where 1 represents genuine note and 0 represents fake note.

Table 1. Dataset description

|  |  |  |
| --- | --- | --- |
| **Attribute Name** | **Value Type** | **Description** |
| Variance of  Wavelet  Transformed Image | Continuous | Variance finds how each pixel varies from the neighbouring pixels and classifies them into different regions. |
| Skewness of Wavelet  Transformed image | Continuous | Skewness is the measure of the lack of symmetry. |
| Kurtosis of  Wavelet  Transformed image | Continuous | Kurtosis is a measure of whether the data are heavy-  tailed or light-tailed relative to a normal distribution. |
| Entropy of image | Continuous | Image entropy is a quantity which is used to describe the  amount of information which must be coded for, by a compression algorithm. |
| Class | Integer | Class contains two values 0 representing genuine note and  1 representing fake note |

**Procedures and Methods**

The methodology began with searching the dataset or gathering data for the identified problem. This step consisted of three steps: identifying various data sources, such as the Kaggle, UCI, and UCSD repositories; collecting data; and finally, integrating the data obtained from various sources. Afterward, data preprocessing is carried out to transform the raw data into a clean data set and carry out operations to fix the problems of missing values, duplication or redundancy, invalid data, and noise in the data set by first importing all the necessary libraries from loading the dataset to performing manipulation and finding and handling the missing values, encoding the categorical data, and also performing feature scaling when and where needed to carry out the Explana In EDA, the analysis of the head, tail, shape, information, description of the data, examination of null values, unique values, and so forth is done, and then graphs and charts, including bar,scatter-plot, heatmap, are used.



The steps to design the proposed system are as follows

1. Loading the libraries and modules
2. Importing Data
3. Finding and handling the missing data
4. Extraction of data
5. Handling categorical data
6. Split the dataset.
7. Feature scaling
8. Visualization
9. Creating the Recommender/Model
10. Run Recommender systems/Models

# 4.0 ALGORITHMS USED

# 4.1 Logistic Regression

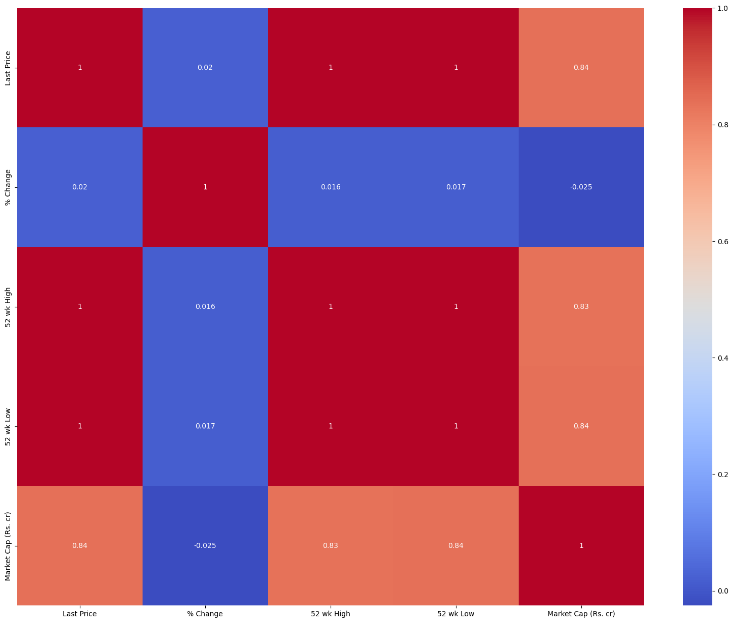
Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables.

Logistic regression predicts the output of a categorical dependent variable. Therefore, the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1.

# 4.2 HEAT MAP

A heat map is a graphical representation of data where values are depicted using color gradients. It provides a visual summary of data by assigning colors to different values, allowing patterns and variations to be easily identified. Heat maps are commonly used in various fields, including statistics, data visualization, finance, and geography.

The typical structure of a heat map consists of a grid or matrix, where each cell represents a data point or a specific location. The intensity or color of each cell corresponds to the magnitude or value of the data it represents. Warmer colors like red or orange often indicate higher values, while cooler colors like blue or green represent lower values.



# 4.3 K-Nearest Neighbor (KNN)

K-Nearest Neighbor is one of the simplest Machine Learning algorithms based on Supervised Learning technique. K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories. K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm.

KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much similar to the new data. It is widely disposable in real-life scenarios since it is non-parametric, meaning, it does not make any underlying assumptions about the distribution of data (as opposed to other algorithms such as GMM, which assume a Gaussian distribution of the given data).

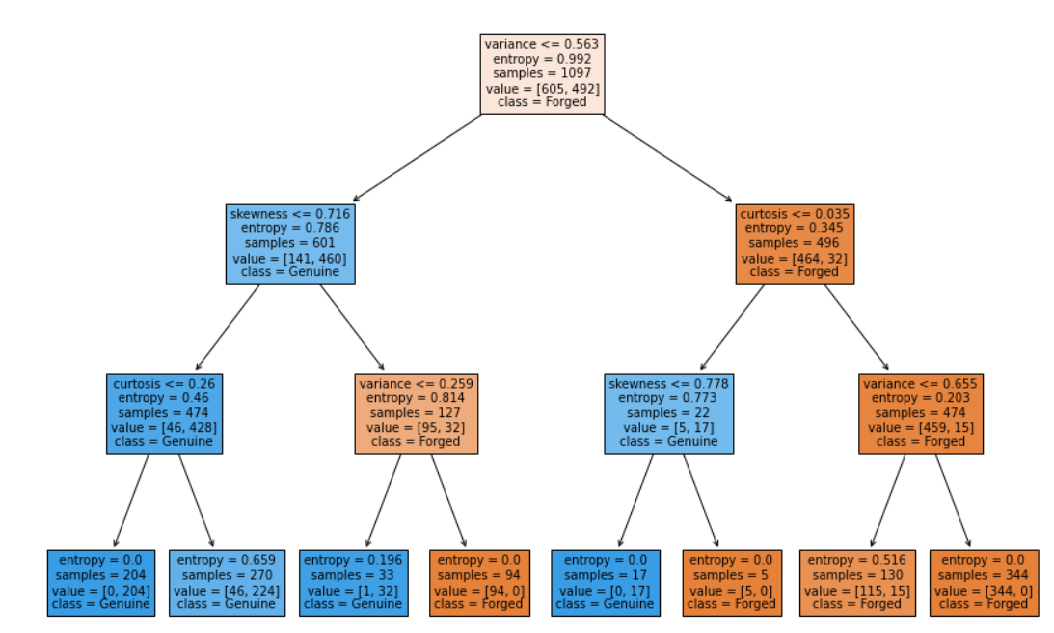


Where D denotes: Euclidean Distance, (X1, X2) and (Y1, Y2) are the two data points

# 4.4 Decision Tree

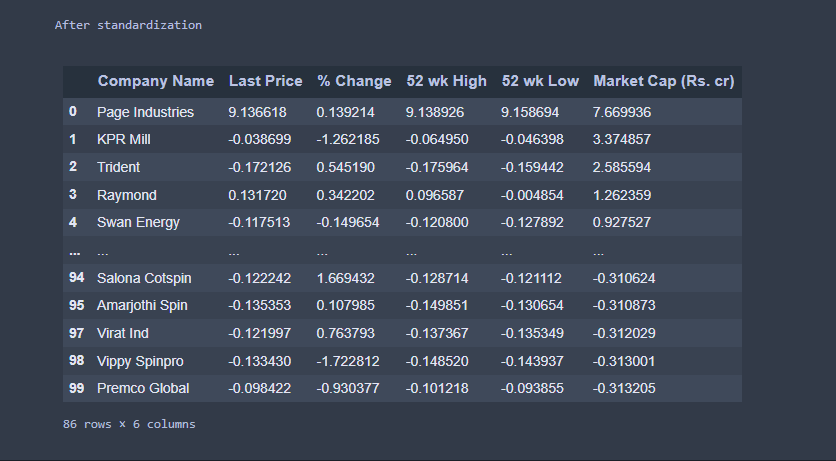
Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome. In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches.

In order to build a tree, we use the CART algorithm, which stands for Classification and Regression Tree algorithm.

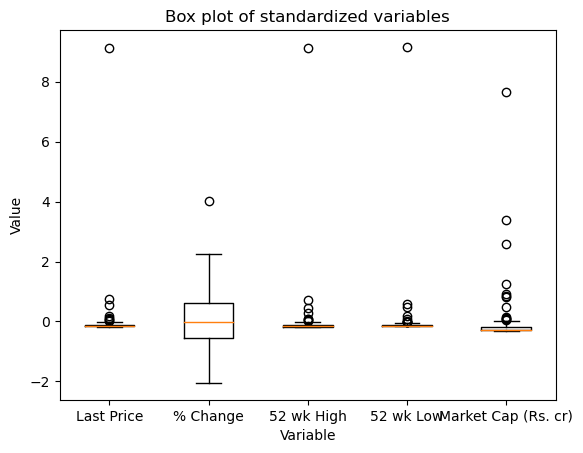


# 4.5 Standardization

Standardization, also known as feature scaling or normalization, is a preprocessing technique commonly used in machine learning regression tasks. It involves transforming the features or variables in a dataset to have a consistent scale or distribution. The goal of standardization is to ensure that all features contribute equally to the learning process and prevent certain features from dominating others due to their original scales



**BOX PLOT VISUAL-**

****

# 5.0 RESULT AND ANALYSIS

# 5.1 Performance Measure

Following measures have been used to measure the performance of the models implemented

 Accuracy – The accuracy of the test is its ability to differentiate the genuine and fake note test cases correctly.

Accuracy = (TP+TN)/(TP+TN+FP+FN)

 Sensitivity - The sensitivity of a test is its ability to determine the genuine note cases correctly.

Sensitivity = TP/(TP+FN)

 Specificity - The specificity of a test is its ability to determine the fake note cases correctly.

Specificity = TN/(TN+FP)

 Precision - The precision of a test is its ability to determine the number of notes that classifier labeled

as genuine is actually genuine

Precision = TP/(TP+FP)

Where:

* + True Positive (TP) = the number of cases correctly identified as genuine notes.
  + True negative (TN) = the number of cases correctly identified as fake notes.
  + False positive (FP) = the number of cases incorrectly identified as genuine notes.
  + False negative (FN) = the number of cases incorrectly identified as fake notes.

# 5.2 Comparative Study

# Hold-out method is used which divides the dataset into the ratio of 80:20 (training data: test data) and following results have been yielded on test data.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Techniques** | **Accuracy** | **Specificity** | **Sensitivity** | **Precision** |
| Logistic Regression | 96.36% | 95.54% | 97.45% | 94.26 |
| Naïve Bayes | 82.54% | 83.43% | 77.11% | 81.25% |
| KNN | 100% | 100% | 100% | 100% |
| Decision Tree  (Criterion = ‘entropy’) | 99.27% | 99.36% | 99.15% | 99.15% |
| Decision Tree  (Criterion = ‘gini’) | 98.90% | 98.08% | 100% | 97.52% |

# 6.0 CONCLUSION

After analyzing various techniques used to detect forged textile stock pricess, this paper presents textile stock prices authentication for recognizing the textile stock prices as genuine or fake by using two supervised learning techniques. Extensive experiments have been performed on textile stock pricess dataset using various models to find the best model suitable for classification of the notes. Performance metrics have been calculated to compare the performances of all the techniques. The result shows that KNN and SVM gives 100% success rate. These techniques are an efficient way of solving the problem for all banking machines that accept all types of notes. In future, this work can be extended by categorizing the notes into different categories as Genuine, Low-Quality forgery, High-Quality forgery.

# REFERENCES

1. Zeinab Shahbazi, Yung-Cheol Byun\* (2020). Product Recommendation Based on Content-based Filtering Using XGBoost Classifier. International Journal of Advanced Science and Technology. 6979 – 6988
2. T. Keerthana1, T. Bhavani2, N. Suma Priya3, V. Sai Prathyusha4, K.Santhi Sri5 (2020). FLIPKART PRODUCT RECOMMENDATION SYSTEM. ISSN NO:0377-9254
3. Haein Kim, Geunho Yang, Hosang Jung, Sang Ho Lee, Jae Joon Ahn(2018). An Intelligent Product Recommendation Model to Reflect the Recent Purchasing Patterns of Customers 163–170
4. Keunho Choi, Donghee Yoo, Gunwoo Kim, Yongmoo Suh. (2012). A hybrid online-product recommendation system: Combining implicit rating-based collaborative filtering and sequential pattern analysis. ECRA, 11, 309-317.
5. Lin Chen, Rui Li, Yige Liu, Ruixuan Zhang, Diane Myung-Kyung Woodbridge. Machine Learning-based Product Recommendation using Apache Spark. Manning publications. 2016
6. Soma Bandyopadhyay, S. S. Thakur1, J. K. Mandal, "Product recommendation for e‐commerce business by applying principal component analysis (PCA) and K‐means clustering: benefit for the society," Springer-Verlag London Ltd., part of Springer Nature 2020.
7. Michael P. O'Mahony, P ́adraig Cunningham and Barry Smyth," An Assessment of Machine Learning Techniques for Review Recommendation Marlin, B. Modeling user rating profiles for collaborative filtering. Advances in Neural Information Processing Systems, 16, 2003, 627–634.
8. A.B.M.Fahim Shahriar, Mahedee Zaman Moon, Hasan Mahmud, Kamrul Hasan, “Online Product Recommendation System by Using Eye Gaze Data”
9. Sanket Hanmant Pawar, Avadhut Prashant Lohar, Omkar Nitin Thopate, Rushikesh Sukhdev Auto, Dr.Y.B.Gurav. “PRODUCT RECOMMENDATION SYSTEM USING MACHINE LEARNING”, || Volume 6 || Issue 4 || April 2021 || ISSN (Online) 2456-0774
10. Image-based recommendations on styles and substitutes J. McAuley, C. Targett, J. Shi, A. van den Hengel SIGIR, 2015
11. Daqing Chen, Sai Liang Sain, and Kun Guo, Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining, Journal of Database Marketing and Customer Strategy Management, Vol. 19, No. 3, pp. 197â€ "208, 2012 (Published online before print: 27 August 2012. doi: 10.1057/dbm.2012.17).s