Science World Journal Vol. 19(No 2) 2024
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University
https://dx.doi.org/10.4314/swj.v19i2.24

MACHINE LEARNING-BASED MODEL FOR PREDICTION OF ACCOUNTANTS BEHAVIOUR IN NIGERIA

*[1]Udeagha E.O., [1]Choji D.N., [2]Olalere M., [2]Ajayi B.A.

[1]Department of Computer Science, University of Jos, Nigeria
[2]Department of Computer Science, Nasarawa State University, Keffi, Nigeria

*Corresponding Author Email Address: emmanjab@gmail.com

ABSTRACT
The integration of technology in accounting roles raises questions about the adaptability and skills of accountants in utilizing these tools effectively. Understanding how accountants' behavior is influenced by technology is crucial for their professional development and the accounting industry's future. This study focused on the development of a predictive model, leveraging both Naive Bayes and K-Nearest Neighbors (KNN) models. The research methodology involved the use of Pandas DataFrame to establish a robust framework for the dataset, incorporating both established and innovative features as input variables. These datasets were then utilized as the training data for the predictive model, with the primary objective of extracting valuable insights for decision-making and forecasting accountant behavior. The key findings of the study shed light on the performance of the different models employed. The Naïve Bayes model emerged as a standout performer, achieving an accuracy rate of 63% and an exceptional recall rate of 97%. This underscores its effectiveness in predicting accountant behavior, especially in identifying positive instances. On the other hand, the K-Nearest Neighbors model displayed a balanced trade-off between precision and recall, achieving an accuracy rate of 52% and an F1 score of 64%. This suggests that the model provides a reasonable compromise between accurately identifying positive cases and overall performance. Furthermore, the hybrid KNN-NB model, which amalgamates elements from both approaches, also achieved an accuracy rate of 52%. This finding indicates that the hybrid model has the potential to harness the strengths of both algorithms, offering a versatile approach to predicting accountant behavior.

Keywords: Machine learning, Predictive Model, Classification, KNN, Naïve Bayes

INTRODUCTION
The digitalization of the world continues and new innovations within accounting and finance will affect every day work tasks. Without insights about what is currently happening on the field and what will most likely change in the foreseeable future, various professions will be put at risk. The gap between information technology and the traditional accounting and finance roles is predicted to rapidly diminish (Huttunen.etal.,2019). The understanding of concepts such as Big Data, Accounting Systems, Machine learning, Cloud Computing, and Data Science Applications will be crucial for succeeding in the coming job market as a decision maker in financial sector. The study focuses on the efficient use of data for business decision-making in the financial industry. The Big Data revolution promises to transform how we live, work, and think by enabling process optimization, empowering insight discovery and improving decision making. The realization of this grand potential relies on the ability to extract value from such massive data through data analytics; machine learning is at its core because of its ability to learn from data and provide data driven insights, decisions, and predictions. However, traditional machine learning approaches were developed in a different era, and thus are based upon multiple assumptions, such as the data set fitting entirely into memory, what unfortunately no longer holds true in this new context (Abdualgalil and Abraham, 2020). These broken assumptions, together with the Big Data characteristics, are creating obstacles for the traditional techniques. Consequently, this thesis compiles, summarizes, and organizes machine learning challenges with Big Data.

The threat that big data analytics will replace many of the tasks traditionally played by accountants may be particularly salient in the auditing context. For example, instead of relying on traditional sampling techniques to perform tests of details, automated processes could examine entire populations for unusual patterns and anomalies. In place of auditors sending out manual confirmations, automatic confirmations could be achieved by a block chain type of technology. A consensus has been emerging among academics that once access to big data analytic techniques becomes ubiquitous in business, users of financial statements will expect audited financial statements on demand, necessitating a shift from traditional sample based auditing to continuous ‘auditing by exception’, where data analytic techniques direct auditor attention on a real-time basis to “instances where the data does not match the auditor’s expectations based on his or her knowledge of the client’s business” (Earley et al., 2018).

Furthermore, it is possible that in the future, public accounting firms may face greater competition for the provision of audit services from non-audit firms. As further advances in data analytic and data visualization techniques become available, it will become easier for non- accountants with competencies in data analysis to obtain audit evidence and complete financial statement audits by applying data analytic techniques to big data (Brown-Liburd et al. 2017; Earley 2018). The use of indigenous developed software for computation of Big Data will greatly enhance the work of accountants, and by so doing give both individuals, corporate organizations, governments at different levels and the public the required data for development which will enrich the society.

Accounting and Finance are no exception to this premise, especially when it comes to predicting the outcome of any business decision. Although decision-making always involves future outcomes, however, the effect of any financial decision could be exceptionally phenomenal on the future of any business. Financial decisions involve day-to-day business operations on the one hand and long-term strategy on the other. A business needs to decide about the pricing of the product on the one hand and the pricing of

https://dx.doi.org/10.4314/swj.v19i2.24

securities on the other. A business needs to review the business risk as well as the financial risk. All these situations involve financial predictions involving a complex interaction between various datasets. Unfortunately, the existing financial theories are not able to handle such complex decision situations, though they may give some idea. (Heaton et al.,2018). According to (Dixon & Halperin 2019) claim that even though machine learning has been prevalent in financial services for over four decades now, only in the past few years, its impact has been felt in investment management and trading. Both computational and theoretical developments in machine learning have resulted in the increased practice of machine learning in the finance area.

METHODOLOGY

This study utilized a research approach termed as design and creation research, a variation of Design Science Research (DSR) commonly applied in the realm of information systems and computing research for addressing issues through a problem-solving framework (Oates, 2006). DSR, grounded in engineering and scientific principles, primarily centers around the development of artifacts (Oates, 2006). Nonetheless, there remains ongoing debate regarding the definition, concepts, terminology, and supporting theories of DSR as a research paradigm (Johannesson et al., 2013). While the tenets of DSR lack a systematic delineation due to the intricate nature of science, information systems, and computing research, it constitutes a field often involving the convergence of organizational, human, and technological elements within problem-solving contexts. Hence, it is particularly well-suited for tackling challenges that encompass organizational, human, and technological dimensions.

Framework for the Model for Predicting Accountant Behaviour

The pseudocode provided is an outline of an algorithm for Predicting Accountant Behavior detection that utilizes naïve bayes and KNN techniques. It describes the basic steps that the algorithm would follow to detect account behavior.

Algorithm for Predicting Accountant Behaviour

1. Import required packages
2. Read dataset
3. Describe data
4. Fill in NaN values
5. Print correlation
6. Extract features
7. Normalization
8. Data splitting
9. Model initializations
10. Model combination
11. Models' evaluation
12. Plots

Formulation of The Dataset

The dataset used in this study was obtained via the administration of questionnaires. The gathered data, as shown in Table 1, underwent a process of cleaning and preprocessing. This dataset provides the researcher with the opportunity to train algorithms using the whole dataset.

From Table 1, the data collected from the survey was then used to train a model, which was used to predict the behavior of accountants using naïve bayes and KNN algorithms. The dataset includes demographic variables (e.g., age, gender, education level), variables related to the use of machine learning in accounting (e.g., familiarity with machine learning, frequency of use), and variables related to attitudes towards machine learning (e.g., perceived usefulness, perceived ease of use). A total of 46 accountants participated in the survey, and the majority of the respondents were males (30%) compared to females (15%). The age distribution of the respondents was between 25 and 55 years, with an average age of 34 years. Furthermore, the highest proportion of respondents had a bachelor's degree (25%), while 10% had a master's degree, and 5% had a PhD. Perceived Benefits of Machine Learning The survey results indicated that the majority of the accountants (35%) believed that machine learning can enhance their job performance. The respondents stated that machine learning could improve their accuracy in financial reporting, reduce the time taken to complete tasks, and increase efficiency in their work processes. Moreover, 10% of the respondents believed that machine learning could help them to develop new skills, while 5% stated that machine learning could enhance their decision-making abilities. The results also revealed that 30% of the accountants perceived challenges in adopting machine learning in their profession. The respondents identified the lack of knowledge and skills as the most significant barrier to the adoption of machine learning in their work. They also cited the cost of acquiring and maintaining machine learning technology and the fear of job displacement as other significant challenges. Furthermore, 10% of the respondents believed that machine learning could pose a threat to the privacy and security of financial data. During this phase it was found that attributes such as no of years of practice, years of view, year of work, awareness, sex and age have a great impact on the mental education, whereas others, like Marital Status and Computer literacy, are considered redundant. The best features according to the results of the previous step were selected and the learning algorithms that had the best accuracy results were run on them.

Performance Evaluation Matrices

Diverse metrics are often used to assess the effectiveness of a categorization model. The measurements include Accuracy, Confusion Matrix, Precision, Recall, F1-score, Error Rate, and Training time.

Classification Accuracy: The term "accuracy" refers to the ratio of correct predictions produced by the model to the total number of predictions made.

$$\text{Accuracy (\%)} = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \qquad (1)$$

Precision: Precision is determined by the ratio of true positives to the total of true positives and false positives.
The formula for precision is:

$$\text{Precision} = \frac{TP}{TP+FP} \qquad (2)$$

Sensitivity; Sensitivity is a numerical metric used to calculate the proportion of correctly identified positive situations that were incorrectly labeled as negative by the model. It is sometimes denoted as recall or true positive rate. Mathematically, it is defined as the ratio of the number of true positive (TP) occurrences to the sum of true positive and false negative (FN) cases.

$$\text{Sensitivity} = \frac{TP}{TP+FN} \qquad (3)$$

 https://dx.doi.org/10.4314/swj.v19i2.24

Specificity
Specificity is a synonym for the actual negative rate. The theoretical definition of the term involves the calculation of the ratio between the number of true negative (TN) instances and the sum of true negative and false positive (FP) cases.
Mathematically, the expression is as follows:

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (4)$$

F-Score
The F-score is a statistical metric used to evaluate the effectiveness of a binary classification model by measuring its capacity to reliably anticipate occurrences of the positive class. The computation employs the metrics of accuracy and recall.
It is mathematically calculated as:

$$\text{F1-score} = 2 * \frac{precision * recall}{precision + recall} \quad (5)$$

Error Rate (EER)
The error rate may be computed by dividing the total count of wrong predictions made on the test set by the total count of predictions made on the test set.
Mathematically, it is expressed as:

$$\text{Error Rate} = \frac{Incorrect\ Predictions}{Total\ Predictions} \quad (32)$$

RESULTS AND DISCUSSION

Numerical Experimental Performance of the Proposed Model
This step evaluates the proposed model's quantitative experimental performance and refines the sample size and feature collection. The findings are thoroughly documented, including tables and figures.

Importing the Libraries
The Python Libraries Pandas, Numpy, CSV, Matplotlib, Seaborn, and Joblib were loaded into Jupyter Notebook. Numpy enables fast computation and broadcasting over multi-dimensional arrays by vectorization (Stančin & Jović, 2019). Pandas is a sophisticated and intuitive open-source program designed for the analysis and manipulation of data. It is constructed using the Python programming language (Subasi, 2020). Matplotlib and Seaborn are Python libraries especially tailored for data visualization. They provide an intuitive interface for generating visually captivating and practical graphs. Seaborn is constructed upon the framework of Matplotlib and provides a slightly smaller range of functionalities in comparison to Matplotlib (Pintor et al., 2019). Figure 1 depicts the inclusion of Pandas, Numpy, CSV, Matplotlib, Seaborn, and Joblib Python libraries into the Jupyter Notebook.

Loading the Dataset
Table 2 depicts the imported dataset. The 'hvplot.pandas' was employed for interactive data visualization, loads a dataset from 'Dataset.csv' using 'pd.read_csv()', and provides initial insights through 'df.head()' and 'df.describe()'. It sets the stage for comprehensive data analysis and decision-making.

Replacing Unknown Values in the Dataset
In the data pre-processing step, the system replaces all columns that contain question marks with non-null values. This process aims to handle missing or unknown values in the dataset. After replacing the question marks with non-null values, the system proceeds to print the updated data. This allows for a visual inspection of the dataset to verify the changes and ensure that the non-null values have been properly substituted. By replacing the question marks with non-null values and printing the data, the system addresses missing or unknown values in the dataset, ensuring that subsequent analysis or modeling can be performed accurately and reliably.

Printing of the Correlation
Table 3 shows the dataset correlation analysis. Eliminating "behavior" and "awareness" columns from the dataset streamlined the prediction process and improved efficiency. After being deemed unnecessary, these columns were removed. Note that deleting these columns made the dataset more concentrated and meaningful (Table 3). This intentional dataset dimensionality reduction has several benefits. First, it makes data analysis easier to handle and analyze. Second, it decreases computing overhead by processing fewer data points, which is especially useful for huge datasets. After removing unneeded columns, the algorithm reassesses variable correlation. After-processing correlation analysis is crucial for understanding the linkages and interdependencies between retained variables. Understanding these connections is crucial to understanding how factors interact and affect prediction. This rigorous technique to data reduction and correlation reevaluation streamlines and refines the dataset. It optimizes the dataset for modelling or analysis, ensuring that the variables are useful to prediction. This increases modeling efficiency, prediction accuracy, and interpretability, resulting in more informed and accurate results.

Feature Extraction
In this step of the data preparation process, the dataset is divided into its fundamental components: features (X) and the target variable (y). This is achieved by utilizing the loaded Pandas DataFrame 'df'. The 'X' variable is formed by excluding two specific columns, "BEHAVIOUR" and "AWARENESS," from the dataset, representing the input attributes for modeling. Simultaneously, the 'y' variable is created by isolating the "BEHAVIOUR" column, representing the target variable that signifies the behavior of interest. This step depicted in depicted in Table 4 is pivotal in organizing the dataset for subsequent data preprocessing, model development, and predictive performance evaluation.

Data Normalization
In this step of the data preparation process, normalization is performed using the 'MinMaxScaler' from the 'scikit-learn' library. This process transforms the feature values to a standardized scale, typically ranging between 0 and 1. Normalization ensures that all features have the same scale, preventing any particular feature from dominating the modeling process due to its larger numerical range. This step enhances the model's ability to effectively learn from the data. The dataset is split into training and testing sets. This is achieved with the 'train_test_split' function, which partitions the data into two subsets. The 'X_train' and 'y_train' sets are designated for training the predictive model, while the 'X_test' and 'y_test' sets are reserved for evaluating the model's performance. The 'test_size' parameter, set at 0.2, indicates that 20% of the data is allocated for testing, leaving 80% for training. Additionally, a 'random_state' is set to ensure reproducibility of the split.

https://dx.doi.org/10.4314/swj.v19i2.24

RESULTS

This section will present the data analysis results based on the model computation, test and comparison of the models and plotting the graph of the computed data set.

Extensive natural introgression is relatively uncommon but interesting when it occurs. From this it is possible to obtain a theoretical insight into the nature of genetic differentiation and of natural selection. Also, the great genetic variability in a hybrid swarm makes a desirable starting point for selective breeding programs. This explains why the Hybrid (Hybrid KNN-NB) performance in the experiment carried out is of low output compared to the NB and KNN outputs.

Model Creation

The Hybrid KNN-NB model proposed in this study combines two distinct models: Naive Bayes (NB) and k-Nearest Neighbors (KNN). This fusion aims to leverage the advantages of both models, resulting in a hybrid approach. This model is aimed at predicting the behavior of accountants in Nigeria based on the use of machine learning.

K-Nearest Neighbors (KNN) Model

The model employed in this study is a k-Nearest Neighbors (k-NN), implemented using the 'KNeighborsClassifier' imported from the 'scikit-learn' library. Figure 4.4 illustrates the model creation process. Subsequently, the k-NN model is initialized with 'n_neighbors' set to 3, signifying that it considers the three nearest data points when making predictions based on features. Following initialization, the 'knn_model' is trained with the provided training data ('X_train' and 'y_train'), enabling it to learn from the patterns in the training dataset. Afterwards, the trained k-NN model is applied to make predictions on the testing data ('X_test'), and these predictions are stored in the 'knn_predictions' variable. Lastly, the 'knn_predictions' are printed, revealing the model's computed predictions based on the input features. This step is pivotal for evaluating the model's performance and its capacity to make accurate predictions on previously unseen test data, serving as a critical aspect of model assessment.

KNN Prediction

Following training, the model is applied to make predictions on the testing data ('X_test'). The predicted values are stored in the 'knn_predictions' variable. Finally, the 'knn_predictions' are printed, showing the model's computed predictions. The output is a binary sequence of 0s and 1s depicted in Figure 2, representing the model's predictions for each corresponding data point in the testing dataset. These predictions are fundamental for evaluating the model's performance and its ability to accurately classify data points into the appropriate categories.

Learning Curve for KNN Model

In Figure 3, a learning curve is presented for the k-Nearest Neighbors (k-NN) model, achieved through the utilization of the 'plot_learning_curve' function. The learning curve aids in making informed decisions about model training, assisting in the diagnosis of issues like overfitting or underfitting, and guiding the optimization of model complexity and data size for enhanced predictive performance.

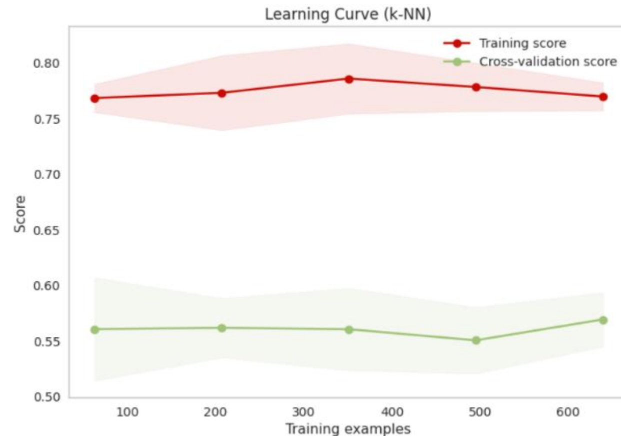

Figure 3: Learning Curve for KNN Model

Naïve Bayes Model

The model utilized in this study is a Naive Bayes model, specifically implemented using the 'Gaussian Naive Bayes (GaussianNB) classifier' imported from the 'scikit-learn' library. The model creation process is illustrated in Figure 4. Following the import, the 'nb_model' is instantiated by creating an instance of the GaussianNB classifier. This particular classifier is well-suited for datasets where the features follow a Gaussian (normal) distribution. Subsequently, the 'nb_model' is trained with the training data, namely 'X_train' and 'y_train.' This training process equips the model with the capability to discern underlying patterns and relationships within the dataset. Following successful training, the trained Naive Bayes model is employed to make predictions on the testing data, which is represented by 'X_test.' The resulting predictions are stored in the 'nb_predictions' variable. Lastly, the 'nb_predictions' are printed, presenting the model's computed predictions based on the input features. This final step is crucial for evaluating the performance of the Naive Bayes classifier and gaining insights into its proficiency in categorizing data points into the appropriate classes. Such insights are valuable for informed decision-making and predictive analysis.

NB

```
In [ ]: from sklearn.naive_bayes import GaussianNB

        # Create Naive Bayes model
        nb_model = GaussianNB()
        nb_model.fit(X_train, y_train)

        # Make predictions
        nb_predictions = nb_model.predict(X_test)
        print(nb_predictions)
```

Figure 4: Naïve Bayes Model

Naïve Bayes Prediction

After the completion of the training process, the trained model is employed to generate predictions on the testing data, denoted as 'X_test,' by utilizing the 'predict' method. The resultant predicted values are stored in the variable named 'nb_predictions.' Subsequently, the 'nb_predictions' are displayed, thereby presenting the model's computed predictions based on the input features. These predictions are represented as a sequence of binary values, specifically 1s and 0s shown in Figure 5, which correspond to the model's classification of each individual data point. This particular step holds significant importance as it serves

Science World Journal Vol. 19(No 2) 2024 https://dx.doi.org/10.4314/swj.v19i2.24
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University

the purpose of evaluating the model's performance and its capability to accurately classify data points into their respective categories. Consequently, this evaluation process aids in facilitating decision-making and predictive analysis.

```
[1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1
 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 0 1 1 1
 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1
 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1]
```

Figure 5: Computed Predictions for Naïve Bayes Prediction Model

Learning Curve for Naïve Bayes Model

In Figure 6, The 'plot_learning_curve' function is employed to create a learning curve for the Naive Bayes model. Learning curves are insightful tools that demonstrate how a model's performance evolves as the size of the training dataset varies. This visualization helps us gauge the model's ability to generalize and identify potential issues like overfitting or underfitting. The 'nb_model,' representing the trained Naive Bayes model, is assessed using this curve. The title of the plot is set as "Learning Curve (Naive Bayes)," with 'X_train' and 'y_train' serving as the training data and labels, respectively. A 5-fold cross-validation approach ('cv=5') is applied, which partitions the dataset into five subsets for robust evaluation.

```
plot_learning_curve(nb_model, "Learning Curve (Naive Bayes)", X_train, y_train, cv=
# plot_residuals(nb_model, X_train, y_train, X_test, y_test)

plt.show()
```

Figure 6: Creating Naïve Bayes Learning Curve

The 'plt.show()' command is used to display the generated plot(s). The generated plot is depicted in Figure 7. These visualizations play a crucial role in assessing and improving the performance of the Naive Bayes model, facilitating informed decision-making in predictive analysis and model refinement.

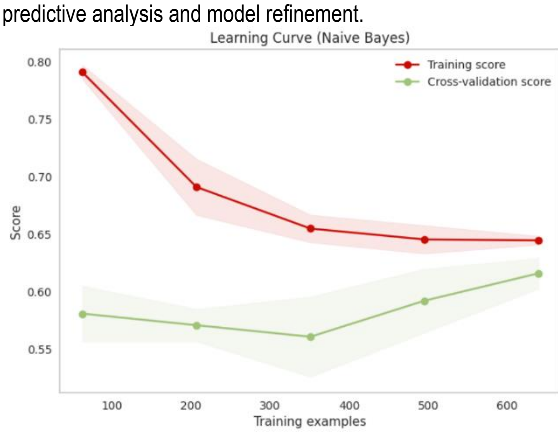

Figure 7: Validation Curve for Naïve Bayes Model

Hybrid KNN-NB

The Hybrid KNN-NB algorithm is a fusion of the K-Nearest Neighbors (KNN) algorithm with the Naive Bayes (NB) classifier. This combination offers a comprehensive methodology for predictive modeling and classification problems. The use of this hybridization method capitalizes on the respective advantages of both algorithms in order to augment the precision and resilience of predictive models. In practical implementation, the procedure entails the preparation of the dataset, followed by the concurrent training of both the K-Nearest Neighbors (KNN) and Naive Bayes (NB) models using the training data. Subsequently, the predictions generated by both models are combined for the test data. The K-Nearest Neighbors (KNN) algorithm demonstrates proficiency in capturing localized patterns, rendering it very useful in handling complicated decision boundaries. On the other hand, the Naive Bayes (NB) algorithm's premise of feature independence has significant value when dealing with datasets that include a large number of dimensions. The hybrid KNN-NB model strives to enhance the accuracy and reliability of predictions by integrating these strengths, hence presenting a viable solution for a wide range of machine learning difficulties. The selection of combination methodology and hyperparameters is customized to the particular issue and dataset, making it a flexible tool within the domain of ensemble learning techniques.

A technique is shown for integrating predictions generated by two distinct machine learning models, namely the K-Nearest Neighbors (KNN) model and the Naive Bayes (NB) model. The purpose of this combination procedure is to provide a forecast that is more resilient and precise by using the respective advantages of both models. The variable 'combined_predictions' is first instantiated to contain the fused predictions. These predictions are obtained by calculating the average of the predictions provided by the KNN and NB models. The objective of this averaging strategy is to get a forecast that is well-balanced by including the results of both models. In order to address instances when the cumulative prediction is precisely 0.5, indicating an equivalent contribution from both models, the code incorporates a procedure to modify these probabilities. In a specific manner, it substitutes these values with random variables that span from 0.1 to 1, so guaranteeing that the aggregate forecast is confined within an acceptable range of probabilities. Following this, the predictions that have been changed are rounded to the nearest whole number, resulting in them being transformed into discrete class labels.

Hybrid KNN-NB Prediction

The variable 'combined_predictions' is first created to hold the aggregated predictions. The predictions are derived via the process of averaging the predictions obtained from both the K-Nearest Neighbors (KNN) and Naive Bayes (NB) models. This approach facilitates the generation of a well-balanced and consensus-based forecast. The 'combined_predictions' are converted to integers by using the 'int(combined_predictions[i])' function. This process transforms the predicted values into distinct categorical labels. The modified and rounded values of the variable 'combined_predictions' are shown in Figure 4.13, which correspond to the ultimate forecasts for the dataset. The predictions are presented as a sequential arrangement of binary values, namely 1s and 0s, as seen in Figure 4.13. Each value represents the classification given by the Hybrid KNN-NB model.

DISCUSSION

Performance Evaluation of Selected Models

In order to choose the most suitable model, an evaluation of performance parameters such as accuracy, precision, recall, and F1 score was conducted on a designated test dataset. The results shown in Table 5 demonstrate the performance of the Naive Bayes (NB) model. The accuracy achieved by the NB model is 63%.

https://dx.doi.org/10.4314/swj.v19i2.24

Additionally, the precision is reported as 64, the recall as 97, and the F-measure as 77. On the other hand, the K-Nearest Neighbors (KNN) model has an accuracy rate of 52%, precision of 61, recall of 68, and an F-measure of 64. The Hybrid NB-KNN model has similarities to the KNN model, achieving an accuracy of 52% and a precision of 61. However, its recall (65) and F-measure (63) are within the range of values seen in the NB and KNN models. These extensive assessments facilitate the process of choosing the best appropriate model for the specific job.

Table 1: Models Performance

Algorithms/Metrics	**Accuracy**	**Precision**	**Recall**	**F-measure**
NB	63	64	97	77
KNN	52	61	68	64
Hybrid KNN-NB	52	61	65	63

After training and evaluating each model on the dataset, we can analyze their results. Based on our analysis, we found that the NB model performed the best in terms of accuracy, precision, recall, and F1 score. It achieved an accuracy of 63%. The results depicted in Figure 8 indicate that the NB algorithm is a more effective method for predicting accountant behavior towards machine learning.

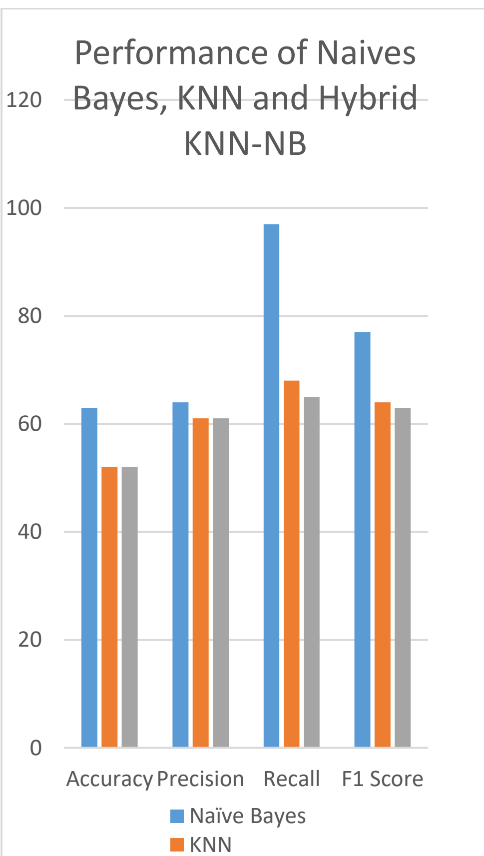

Figure 8: Eight ML Models Accuracy Plot

CONCLUSION

This paper undertakes a thorough investigation of the dynamic terrain where technology, accounting, and human behavior converge. The research primarily focuses on examining the complex interplay between accountants and the disruptive influences of information technology. It extensively investigates the effects of many topics, such as Big Data, Accounting Systems, Machine Learning, Cloud Computing, and Data Science, on the field of accounting. The study utilizes a methodological framework that integrates elements of positivism with interpretivism, enabling the researcher to conduct empirical investigations while also gaining a comprehensive grasp of the research topic.

A crucial element of this research is on the formulation and implementation of prognostic models aimed at comprehending and foretelling the conduct of accountants within the framework of contemporary technology. The concept utilizes Python, a highly adaptable programming language well recognized for its proficiency in data analysis and Machine Learning (ML). The use of Python is a deliberate decision due to its strategic advantages, as it provides users the ability to utilize robust libraries such as NumPy and Pandas for efficient data processing, as well as Matplotlib and Seaborn for effective data visualization. The use of a strong technical infrastructure guarantees that the study is firmly rooted in rigorous data analysis methodologies.

The core of this research is on the development and execution of predictive models, specifically emphasizing the K-Nearest Neighbors (KNN) and Naïve Bayes (NB) algorithms. These models are used both alone and in combination to form a hybrid model known as KNN-NB. The objective of this study is to evaluate and contrast the efficacy of various models in forecasting the behavior of accountants. The study utilizes assessment measures, including accuracy, precision, recall, and F1 score, to assess the efficacy of each model. The results of this research provide significant perspectives on the changing responsibilities of accountants in light of technology progress. The Naïve Bayes model has great performance, obtaining an accuracy rate of 63% and displaying a notable capability in accurately identifying positive cases, as shown by its high recall of 97%. In contrast, the K-Nearest Neighbors model demonstrates a balanced precision and recall, yielding an F1 score of 64%, while obtaining an accuracy rate of 52%. The hybrid KNN-NB model, which integrates components from both the K-Nearest Neighbors (KNN) and Naive Bayes (NB) techniques, has a 52% accuracy rate, suggesting its capability in leveraging the respective advantages of both algorithms. This study serves as a fundamental basis for future inquiries into the modeling and simulation of accountant behavior within the context of artificial intelligence and data analytics. The paper recommends that accountants and accounting firms should continually improve their knowledge regarding machine learning/ artificial intelligence as this will enhance the performance of accounting functions, thereby eliminating certain accounting cost.

Conflict of Interest: The corresponding author, representing all authors, confirms the absence of any conflict of interest.

APPENDIXES

Science World Journal Vol. 19(No 2) 2024
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University
https://dx.doi.org/10.4314/swj.v19i2.24

Table 2: Brief Demography of Participants

	A	B	C	D	E	F	G	H	I
1	S/N	AGE	NO OF YEAR OF PRACTICE	YEARS OF VIEW	YEAR OF WORK	AWARENESS	YEARS OF AWARENESS	% OF COMPUTER LITERACY	NO OF YEARS
2	1	35	5	2.5	8	1	23	60	25
3	2	35	5	2.5	2.5	1	2.5	40	5
4	3	45	10	13	13	0	8	60	10
5	4	35	5	8	2.5	1	2.5	40	10
6	5	45	10	13	13	0	8	20	10
7	6	55	20	13	13	1	8	60	10
8	7	55	20	18	18	0	13	60	15
9	8	45	10	13	18	0	8	60	10
10	9	45	10	13	8	0	13	40	15
11	10	45	10	13	13	0	13	60	20
12	11	45	10	8	8	0	8	60	5
13	12	55	10	18	18	0	18	60	10
14	13	45	20	13	13	1	2.5	60	15
15	14	45	20	8	8	0	13	20	5
16	15	55	10	18	13	0	13	80	15
17	16	45	10	8	8	0	8	80	15
18	17	45	10	2.5	13	0	13	80	20
19	18	45	5	8	8	1	13	60	15
20	19	35	10	2.5	2.5	1	2.5	80	15

Table 3: Imported dataset and headers

	AGE	NO OF YEAR OF PRACTICE	YEARS OF VIEW	YEAR OF WORK	AWARENESS	YEARS OF AWARENESS	PER_OF_COMPUTER_LITERACY	NO_OF_YEARS	TRAIN
count	999.000000	999.000000	999.000000	999.000000	999.000000	999.000000	999.000000	999.000000	999.00
mean	40.362362	12.333333	10.260761	9.799299	0.496496	12.608108	50.554555	14.867868	70.45
std	8.826020	4.563952	5.006791	4.861046	0.500238	6.454079	17.267806	6.088900	18.13
min	25.000000	5.000000	2.000000	2.000000	0.000000	2.000000	0.000000	5.000000	40.00
25%	33.000000	8.000000	6.000000	6.000000	0.000000	7.000000	36.000000	10.000000	55.00
50%	41.000000	12.000000	10.000000	10.000000	0.000000	12.000000	51.000000	15.000000	70.00
75%	48.000000	16.000000	15.000000	14.000000	1.000000	18.000000	65.000000	20.000000	86.00
max	55.000000	20.000000	18.000000	18.000000	1.000000	23.000000	80.000000	25.000000	100.00

	MENTAL_EDUCATION	BIG_DATA_USE	POSITIVE_EFFECT	THREAT	SKILLS	SKILLS_2	LEVEL_OF_WORK	TIME_OF_WORK	RESULT	L
count	999.000000	999.000000	999.000000	999.000000	999.000000	999.000000	999.000000	999.000000	999.000000	999.0
mean	39.179179	79.568569	51.487487	49.496496	68.840841	51.438438	49.870871	3.479980	6.086386	59.73
std	23.757225	12.504631	24.259486	24.065453	18.062120	13.093389	11.855017	1.116285	2.046194	17.81
min	0.000000	20.000000	10.000000	10.000000	10.000000	30.000000	30.000000	2.000000	1.500000	10.0
25%	18.000000	69.000000	30.000000	29.000000	54.000000	40.000000	40.000000	2.750000	4.000000	44.0
50%	40.000000	80.000000	52.000000	50.000000	69.000000	52.000000	50.000000	3.000000	6.000000	61.0
75%	60.000000	90.000000	73.000000	70.000000	84.000000	61.000000	60.000000	4.000000	8.000000	75.0
max	80.000000	100.000000	100.000000	90.000000	100.000000	100.000000	70.000000	9.000000	9.500000	90.0

Table 4: Reduced Data after Dropping Irrelevant Columns

	AGE	NO OF YEAR OF PRACTICE	YEARS OF VIEW	YEAR OF WORK	AWARENESS	YEARS OF AWARENESS	PER_OF_COMPUTER_LITERACY	NO_OF_YEARS
AGE	1.000000	0.043714	0.006011	0.016866	-0.041698	0.035310	-0.010366	0.017300
NO OF YEAR OF PRACTICE	0.043714	1.000000	0.022502	-0.036184	-0.033063	0.023404	-0.005476	0.038112
YEARS OF VIEW	0.006011	0.022502	1.000000	0.035335	0.015868	0.038275	-0.013067	0.002265
YEAR OF WORK	0.016866	-0.036184	0.035335	1.000000	-0.012342	0.021947	0.043298	0.027303
AWARENESS	-0.041698	-0.033063	0.015868	-0.012342	1.000000	-0.032315	0.025745	0.000177
YEARS OF AWARENESS	0.035310	0.023404	0.038275	0.021947	-0.032315	1.000000	-0.007147	0.030234
PER_OF_COMPUTER_LITERACY	-0.010366	-0.005476	-0.013067	0.043298	0.025745	-0.007147	1.000000	0.028878
NO_OF_YEARS	0.017300	0.038112	0.002265	0.027303	0.000177	0.030234	0.028878	1.000000
TRAINING	0.009832	0.007914	-0.041744	-0.024927	-0.017609	-0.069110	0.038516	0.014121
MENTAL_EDUCATION	0.037375	-0.018396	0.002989	0.048601	-0.022248	0.031189	0.042079	0.017931
BIG_DATA_USE	0.005830	0.004840	-0.014318	0.019262	0.004003	0.007587	-0.052521	-0.029715
POSITIVE_EFFECT	-0.003760	-0.062339	-0.016082	0.061196	0.011081	-0.055553	0.002052	-0.106687
THREAT	-0.015689	-0.002603	0.002438	-0.021490	0.004972	0.049286	0.030251	0.073459
SKILLS	0.007383	0.058649	-0.007568	0.005028	0.019290	0.028642	0.019219	0.045135
SKILLS_2	-0.051796	-0.043546	-0.005575	-0.031936	0.003753	0.008136	0.012201	-0.018339

	MENTAL_EDUCATION	BIG_DATA_USE	POSITIVE_EFFECT	THREAT	SKILLS	SKILLS_2	LEVEL_OF_WORK	TIME_OF_WORK	RESULT	LEVEL
AGE	0.037375	0.005830	-0.003760	-0.015689	0.007383	-0.051796	0.078246	0.032010	0.012829	0.013524
NO OF YEAR OF PRACTICE	-0.018396	0.004840	-0.062339	-0.002603	0.058649	-0.043546	0.041817	-0.043039	-0.018119	-0.025759
YEARS OF VIEW	0.002989	-0.014318	-0.016082	0.002438	-0.007568	-0.005575	-0.002521	0.017922	0.044565	-0.018820
YEAR OF WORK	0.048601	0.019262	0.061196	-0.021490	0.005028	-0.031936	-0.010970	-0.008728	0.004278	-0.047468
AWARENESS	-0.022248	0.004003	0.011081	0.004972	0.019290	0.003753	-0.009792	-0.012686	-0.025498	-0.029054
YEARS OF AWARENESS	0.031189	0.007587	-0.055553	0.049286	0.028642	0.008136	0.063115	-0.021639	-0.065674	0.036389
PER_OF_COMPUTER_LITERACY	0.042079	-0.052521	0.002052	0.030251	0.019219	0.012201	0.058696	0.021837	-0.041088	0.001113
NO_OF_YEARS	0.017931	-0.029715	-0.106687	0.073459	0.045135	-0.018339	0.041574	0.006392	0.007391	0.028498
TRAINING	0.088672	0.039580	-0.007331	-0.042449	-0.039623	0.061417	-0.029520	-0.036725	0.000129	0.012829
MENTAL_EDUCATION	1.000000	0.068342	-0.004182	-0.037635	-0.007163	-0.013653	0.022638	0.049820	0.016157	0.000628
BIG_DATA_USE	0.068342	1.000000	0.015716	0.022086	0.036305	-0.028170	-0.028893	-0.047889	0.014499	0.024667
POSITIVE_EFFECT	-0.004182	0.015716	1.000000	0.014398	-0.091124	0.118924	-0.071204	0.061116	0.004070	-0.001701
THREAT	-0.037635	0.022086	0.014398	1.000000	-0.014712	0.025108	-0.051762	-0.046272	0.037226	0.030030
SKILLS	-0.007163	0.036305	-0.091124	-0.014712	1.000000	-0.064911	0.009066	0.063354	-0.029659	0.081992
SKILLS_2	-0.013653	-0.028170	0.118924	0.025108	-0.064911	1.000000	0.001133	0.009616	0.044015	-0.029508
LEVEL_OF_WORK	0.022638	-0.028893	-0.071204	-0.051762	0.009066	0.001133	1.000000	-0.003944	-0.013295	0.026657

Table 5: Features Extraction

```
[0. 0. 1. 1. 1. 1. 1. 0. 1. 1. 0. 1. 1. 1. 0. 1. 0. 1. 1. 1. 1. 1. 1. 1.
 1. 0. 0. 0. 1. 0. 0. 1. 0. 1. 1. 0. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 0.
 1. 0. 1. 1. 1. 1. 1. 1. 0. 0. 1. 1. 0. 1. 1. 1. 1. 1. 1. 0. 1. 0. 0. 1.
 0. 1. 0. 1. 1. 0. 0. 0. 1. 1. 0. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1.
 1. 1. 0. 1. 1. 1. 1. 0. 1. 0. 0. 1. 1. 1. 1. 1. 0. 1. 1. 0. 0. 0. 0. 1.
 1. 1. 1. 1. 1. 0. 0. 0. 1. 1. 0. 1. 1. 0. 0. 1. 0. 1. 1. 1. 1. 1. 1. 0.
 1. 1. 1. 0. 1. 0. 0. 1. 1. 1. 0. 0. 1. 1. 0. 1. 0. 1. 0. 1. 1. 0. 0. 1.
 0. 1. 0. 1. 1. 0. 1. 0. 1. 1. 1. 1. 0. 1. 1. 0. 1. 0. 1. 1. 1. 1. 1. 1.
 0. 0. 1. 1. 1. 1. 1. 0.]
```

Figure 4. 1: Computed Predictions for Hybrid KNN-NB

```
     MENTAL_EDUCATION  BIG_DATA_USE  POSITIVE_EFFECT  THREAT  SKILLS  \
0                  20           100              100      90      90
1                   0            60               80      50      50
2                  40           100              100      10      50
3                  40            80               80      10      30
4                  40            40               60      50      70
..                ...           ...              ...     ...     ...
994                24            98               25      44      69
995                33            75               87      69      41
996                27            72               13      69      65
997                65            98               32      89      78
998                67            61               69      34      63

     SKILLS_2  LEVEL_OF_WORK  TIME_OF_WORK  RESULT  LEVEL
0         100             60           2.5     9.5     50
1          40             30           5.5     7.5     30
2          40             50           5.5     5.5     50
3          60             30           5.5     7.5     30
4          40             30           5.5     5.5     50
..        ...            ...           ...     ...    ...
994        43             59           2.0     4.0     74
995        66             35           2.0     3.0     82
996        56             58           3.0     4.0     88
997        66             65           2.0     3.0     54
998        44             47           2.0     6.0     72
```

```python
import pandas as pd  # For data manipulation
import numpy as np  # For scientific computing
import csv
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.graph_objects as go
import joblib
```

Figure 1: Importing Python Libraries

```
[0 0 1 1 1 1 1 0 1 1 0 1 1 1 0 1 0 1 1 1 1 1 1 1 0 0 1 1 0 0 1 0 1 1 0 1
 1 1 1 1 1 1 1 1 1 1 0 1 0 1 1 1 1 1 1 0 1 1 0 1 1 1 1 1 0 1 0 0 1 0 1
 0 1 1 0 0 0 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 0 1 0 0 1 1 1 1
 1 0 1 1 0 0 1 0 1 1 1 1 1 0 0 0 1 1 0 1 1 0 0 1 0 1 1 1 1 1 1 1 1 1 1 0
 1 0 0 1 1 1 0 0 1 1 0 1 0 1 0 1 1 0 0 1 0 1 0 1 1 0 1 1 1 1 1 1 0 1 1 0 1
 0 1 1 1 1 1 1 0 0 1 1 1 1 1 0]
```

Figure 2: Computed Predictions for K-Nearest Neighbors (KNN) Model

REFERENCE

Abd-Mutalib, H., Muhammad Jamil, C. Z., Mohamed, R., & Ismail, S. N. A. (2023). The determinants of environmental knowledge sharing behavior among accounting educators: a modified theory of planned behavior. International Journal of Sustainability in Higher Education, 24(5), 1105-1135.

Abdualgalil and Abraham (2020): Naive Bayes Modest Probabilistic Classifier Based on the Bayesian: A Review. International Journal of Advanced Research in Computer and Communication Engineering, 2(1). 17-21

Abdural, K. K., Jha, R., & Afroz, S. (2020). Data Mining Techniques for Intrusion Detection: A Review. International Journal of Advanced Research in Computer and Communication Engineering, 3(6). 19-23

Agarwal, R., Joshi, M.V.:(2017): PNrule: A New Framework for Learning Classifier Models in Data Mining. Tech. Report, Dept. of Computer Science, University of Minnesota

Aksoy, S.: k-Nearest Neighbor Classifier and Distance Functions. Technical Report, Department of Computer Engineering, Bilkent University (February 2016)

Alcazar, et al. (2020). Classical Versus Quantum Models in Machine Learning: Insights from a Finance Application. Machine Learning: Science and Technology, 1(3), 035003 https://doi.org/10.1088/2632-2153/ab9009American Journal of Industrial and Business Management, 8, 1817-1824. https://doi.org/10.4236/ajibm.2018.88123

Arnott, R., Harvey, C. R., & Markowitz, H. (2019). A Backtesting Protocol in the Era of Machine Learning. The Journal of Financial Data Science, 1(1), 64-74.

Aziz, S., Dowling, M. M., Hammami, H., & Piepenbrink, A. (2019). Machine Learning in Finance: A Topic Modeling Approach. http://dx.doi.org/10.2139/ssrn.3327277

Azizam, N. A., Dzulkipli, M. R., Shamimi, N. I., Maon, S. N., John, D., Belawing, J., ... & Yahaya, N. (2020). Applicability of theory of planned behavior and protection motivation theory in predicting intention to purchase health insurance. Advances in Business Research International Journal, 6(1), 1-9.

Azman, N. L. A., & Vaicondam, Y. (2020). Behavioral intention in forensic accounting services. International Journal of Psychosocial Rehabilitation, 24(2), 1837-1846.

Bhargava, N., Sharma, G., Bhargava, R., & Mathuria, M. (2018). Decision tree analysis on j48 algorithm for data mining. Proceedings of International Journal of Advanced Research in Computer Science and Software Engineering, 3(6).

Bhavsar, Y. B., & Waghmare, K. C. (2019). Intrusion detection system using data mining technique: Support vector machine. International Journal of Emerging Technology and Advanced Engineering, 3(3), 581-586

Bracke, P., & Datta, et al. (2019). Machine Learning Explain ability in Finance: An Application to Default Risk Analysis. Bank of England Working Paper No. 816. http://dx.doi.org/10.2139/ssrn.3435104

Brown, J. O., Hays, J., & Stuebs Jr, M. T. (2016). Modeling accountant whistleblowing intentions: Applying the theory of planned behavior and the fraud triangle. Accounting and the Public Interest, 16(1), 28-56.

Chukwuani, V. N., & Egiyi, M. A. (2020). Automation of Accounting Processes: Impact of Artificial Intelligence. International Journal of Research and Innovation in Social Science (IJRISS), 4, 444-449.

Chukwudi, O., Echefu, S., Boniface, U., & Victoria, C. (2018). Effect of Artificial Intelligence on the Performance of Accounting Operations among Accounting Firms in South East Nigeria. Asian Journal of Economics, Business and Accounting, 7, 1-11. https://doi.org/10.9734/AJEBA/2018/41641

Conner, M. (2020). Theory of planned behavior. Handbook of sport psychology, 1-18.

Conner, M., & Armitage, C. J. (1998). Extending the theory of planned behavior: A review and avenues for further research. Journal of applied social psychology, 28(15), 1429-1464.

Dixon, M., & Halperin, I. (2019). The Four Horsemen of Machine Learning in Finance.

Economia.https://www.icaew.com/technical/technology/artificial-intelligence/artificial-intelligence-articles/how-artificial-intelligence-will-impact-accounting

Eleonora P.S. (2018). How artificial intelligence is challenging accounting profession. Journal of International Scientific Publication. Vol 12. pp.126-140

Emerson, et al. (2019). Trends and Applications of Machine Learning in Quantitative Finance. 8th International Conference on Economics and Finance Research (ICEFR 2019). Retrieved fromhttps://ssrn.com/abstract=3397005

Farid, B. (Ed.). (2010). Handbook of social network technologies and applications. Springer Science & Business Media.

Fray and Osband (2013): k-Nearest Neighbor Classifier and Distance Functions. Proceedings of International Journal of Advanced Research in Computer Science and Software Engineering, 3(5), 271-278

Frey, G. P. C., Yu, J. X., Yu, P. S., & Lu, H. (2016, August). Parameter free bursty events detection in text streams. In Proceedings of the 31st international conference on Very large data bases (pp. 181-192). VLDB Endowment

Ganti, M., Prasad, P., & Wankhade, L. K. (2021, September). A new data mining based network intrusion detection model. In Computer and Communication Technology (ICCCT), 2021International Conference on (pp. 731-735). IEEE.

Gogas, P., & Papadimitriou, T. (2021). Machine Learning in Economics & Finance. Computational Economics, 57, 1-4. https://doi.org/10.1007/s10614-021-10094-w

Grano, C., Singh Solorzano, C., & Di Pucchio, A. (2022). Predictors of protective behaviours during the Italian COVID-19 pandemic: An application of protection motivation theory. Psychology & Health, 37(12), 1584-1604.

Griffin, O. (2019, October 6). How Artificial Intelligence Will Impact Accounting.

Grindley, E. J., Zizzi, S. J., & Nasypany, A. M. (2008). Use of protection motivation theory, affect, and barriers to understand and predict adherence to outpatient rehabilitation. Physical therapy, 88(12), 1529-1540.

Heaton, J. B., Polson, N. G., & Witte, J. H. (2017). Deep Learning in Finance: Deep Portfolios. Applied Stochastic Models in Business and Industry, 3(1), 3-12. https://doi.org/10.1002/asmb.2209

Hetal Patel and Dharmendra Patel. (2014). "A Brief Survey of Data Mining Techniques Applied to Agricultural Data", International Journal of Computer Application, Vol. 95 (9) 6-

https://doi.org/10.1016/j.procs.2019.04.171

https://doi.org/10.1109/COMITCon.2019.8862214

https://www.rsisinternational.org/journals/ijriss/Digital-Library/volume-4-issue-8/444-449.pdf

Huang, et al. (2020). Deep learning in finance and banking: A literature Review and classification. Frontiers of Business Research in China, 14, 2-24. https://doi.org/10.1186/s11782-020-00082-6

Huang, Z. (2018). Discussion on the Development of Artificial Intelligence in Taxation.

Hulse, J.V., Khoshgoftaar, T.M., Napolitano, A.: Experimental Perspectives on Learning from Imbalanced Data. Appearing in Proceedings of the 24th International Conference on Machine Learning, Corvallis, OR (2007)

Huttunen. Jeong, K., Hong, T., Chae, M. and Kim, J. (2019) Development of a Decision Support Model for Determining the Target Multi-Family Housing Complex for Green Remodeling Using Data Mining Techniques. Energy and Buildings, 202, Article ID: 109401. https://doi.org/10.1016/j.enbuild.2019.109401

ICEAW (2018). Artificial intelligence and the future of accountancy. ISBN 978-1-78363-816-1. Available: www. icaew.com/itfac. (Accessed 25 July 2021)

Israel, et al. (2020). Can Machines 'Learn' Finance? Journal of Investment Management. Available at https://doi.org/10.37200/IJIM/ RP200125

Jalota, C. and Agrawal, R. (2019) Analysis of Educational Data Mining using Classification. International Conference on Machine Learning, Big Data, Cloud and Parallel Computing COMITCon), 14-16:

Kaur N., Sahdev S. L., Sharma M. & Siddiqui L. (2020). Banking 4.0: "the influence of artificial intelligence on the banking industry & how AI is changing the face of modern-day banks". International Journal of Management. Vol. 11 No. 6. pp. 577-585.

Kaur, B., Ahuja, L. and Kumar, V. (2019) Crime against Women: Analysis and Prediction Using Data Mining Techniques. International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), 14-16 February 2019, Faridabad. https://doi.org/10.1109/COMITCon.2019.8862195

Kaur, S. and Bawa, R.K. (2017) Data Mining for diagnosis in Healthcare Sector-a review, International Journal of Advances in Scientific Research and Engineering.,1 (1), 51-59

Keeney C. (2018). How natural language generation drives business outcomes. Available:https://narrativescience.com/Resources/Resource-Library/ArticleDetail-Page/how natural language-generation-drives-business outcomes (accessed 13 July 2021)

Kefiyalew B. B. (2021). Exploring current opportunity and threats of artificial intelligence on small and medium enterprise accounting function; evidence from south west part of ethiopia, oromiya, jimma and snnpr, bonga. Academy of Accounting and Financial Studies Journal. Vol. 25 No. 2.

Kolm, P. N., & Ritter, G. (2019). Modern Perspectives on Reinforcement Learning in Finance. The Journal of Machine Learning in Finance, 1(1). http://dx.doi.org/10.2139/ssrn.3449401

Kumar Doshi, H. A., Balasingam, S., & Arumugam, D. (2020). Artificial Intelligence as a Paradoxical Digital Disruptor in the Accounting Profession: An Empirical Study amongst Accountants. International Journal of Psychosocial Rehabilitation, 24,873-885. https://doi.org/10.37200/IJPR/V24I2/PR200396

Lee, C. S., & Tajudeen, F. P. (2020). Usage and Impact of Artificial Intelligence on Accounting: 213 Evidence from Malaysian Organizations. Asian Journal of Business and Accounting, 13, 213-240. https://doi.org/10.22452/ajba.vol13no1.8

Leitner-Hanetseder, S., Lehner, O.M., Eisl, C. & Forstenlechner, C. (2021), "A profession in transition: actors, tasks and roles in AI-based accounting", Journal of Applied Accounting Research Vol. 22 No. 3. pp. 539-556

Li, Y., Li, Z., & Wang, R. (2011, September). Intrusion detection algorithm based on semi-supervised learning. In Information Technology, Computer Engineering and Management Sciences (ICM), 2011 International Conference on (Vol. 2, pp. 153-156). IEEE

Liu W. et al. (2017). A Survey of Deep Neural Network Architectures and their Applications. Neuro Computing,234, 11-26.

Lopez de Prado, M. (2016). Mathematics and economics: A Reality Check. The Journal of Portfolio Management, 43(1), 1-4. https://doi.org/10.3905/jpm.2016.43.1.001

Luan, H., Geczy, P., Lai, H., Gobert, J., Yang, S. J. H., Ogata, H., Baltes, J., Guerra, R., LiP., & Tsai, C.-C. (2020). Challenges and Future Directions of Big Data and Artificial Intelligence in Education. Frontiers in Psychology, 11, Article ID: 580820https://doi.org/10.3389/fpsyg.2020.580820

Masrom, S., Tarmizi, M. A., Halid, S., Rahman, R. A., & Abd, A. S. (2023). Machine learning in predicting anti-money laundering compliance with protection motivation theory among professional accountants. International Journal of Advanced and Applied Sciences, 10(7), 48-53.

Max G., Dan K., Okyanus P., Conelius S., Minna M. & Othmar L. (2020). Current challenges in implementation of smart robotic process automation in accounting and auditing. ACRN Journal of Finance and Risk Perspective. 9(2020) pp.99-102

Mohammad, S. J. et al. (2020). How Artificial Intelligence Changes the Future of Accounting Industry. International Journal of Economics and Business Administration, 8,478-488. https://doi.org/10.35808/ijeba/538

Mullainathan, S., & Jann, S. (2017). Machine Learning: An Applied Econometric Approach. Journal of Economic Perspectives, 31(2), 87-106. https://doi.org/10.1257/jep.31.2.87

Najjar D. (2019). Is artificial intelligence (AI) the future of accounting? The Balance Small Business. Available: https://www.thebalancesmb.com/is-artificial-intelligence-thefuture-of-accounting-4083182 (accessed 13 July 2021)

Olaiya, F and Adeyemo, A and Barnabas (2012). "Application of Data Mining Techniques in Weather Prediction and Climate Change Studies". International. Journal of Information Engineering and Electronic Business, 2012, 1, 51-59

Özer, G., & Yilmaz, E. (2011). Comparison of the theory of reasoned action and the theory of planned behavior: An application on accountants' information technology usage. African Journal of Business Management, 5(1), 50-58.

Quoc, T. N., Bao, Q. P. T., Huu, B. N., & Bao, A. N. P. (2023, May). Motivating Accounting Information Systems Security Policy Compliance: Insight from the Protection Motivation Theory and the Theory of Reasoned Action. In International Conference on Emerging Challenges: Strategic Adaptation in The World of Uncertainties (ICECH 2022) (pp. 342-359). Atlantis Press.

Reddy, P. S., Yasaswi, K. R. K., & Kumar, B. K. (2019). Accounting Intelligence—TheNew Era in Accounting. Journal of Information and Computational Science, 9,692-697. http://www.joics.org/gallery/ics-1569.pdf

https://dx.doi.org/10.4314/swj.v19i2.24

Renault, T. (nd.). (2020). Sentiment Analysis and Machine Learning in Finance: A Comparison of Methods and Models on One Million Messages. Digit Finance, 2, 1-13.

RománOrús, et al. (2019). Quantum computing for finance: Overview and prospects. Reviews in Physics, 4, 100028, https://doi.org/10.1016/j.revip.2019.100028

Rundo, F. et al. (2019). "Machine Learning for Quantitative Finance Applications: A Survey. Appl. Sci., 9(24)

Shaffer, K.J., Gaumer, C.J. and Bradley, K.P. (2020), "Artificial intelligence products reshape accounting: time to re-train", Development and Learning in Organizations, 34(6), pp. 41-43. https://doi.org/10.1108/DLO-10-2019-0242 (accessed 25 July 2021)

Shavlik, J. & Dietterich, T. (1990). Readings in Machine Learning. San Mateo, CA: Morgan Kaufmann.

Silva, J., Borré, J.R., Castillo, A.P.P., Castro, L. and Varela, N. (2019) Integration of Data Mining Classification Techniques and Ensemble Learning for Predicting the Export Potential Science, 151, 1194-1200.

Suleiman J.M., Amneh K.H., Hela B., Phung A.T., Muhammad S.S. & Ali A.A. (2020). How artificial intelligence changes the future of accounting industry. International Journal of Economics and Business Administration. Vol. III. No. 3. pp.478-488.

Ucoglu, D. (2020). Current Machine Learning Applications in Accounting and Auditing. Press academia, 12, 1-7. https://doi.org/10.17261/Pressacademia.2020.1337

Ukpong, E. G., Udoh, I. I., & Essien, I. T. (2019). Artificial Intelligence: Opportunities, Issues and Applications in Banking, Accounting, and Auditing in Nigeria. Asian Journal of Economics, Business and Accounting, 10, 1-6. https://doi.org/10.9734/ajeba/2019/v10i130099

Wagner, W. P. (2017). Trends in Expert System Development: A Longitudinal Content Analysis of over Thirty Years of Expert System Case Studies. Expert Systems with Applications, 76, 85-96.

Warkula, Y. Z., & Pontoh, G. T. (2023). BEHAVIORAL ACCOUNTING: THE BEHAVIORAL IMPACT OF FISHERIES SOCIAL ASSISTANCE RECIPIENTS OF ARU ISLANDS REGENCY WITH A PLANNED BEHAVIOR THEORY APPROACH. Jurnal Riset Akuntansi dan Bisnis Airlangga Vol, 8(1), 1418-1434.

Windmann, A. (2020). Machine Learning in Finance. University of RWTH Aachen.

Xu, X.: Adaptiv Intrusion Detection Based on Machine Learning: Feature Extracton, Classifier Construction and Sequential Pattern Prediction. International Journal of Web Services Practices 2(1-2), 49–58 (2006)

Yanling S. (200). The impact of artificial intelligence on the accouning industry. Retrieved from https://doi.org/10.1007/978- 3-030-15235-2_129. (accessed 10 July 2021)

Yeung, D.Y.,Chow, C.: Prazen-window Network Intrusion Detctors. In: 16th International Conference on Pattern Recognition, Quebec, Canada, pp. 11–15 (August 2002)

Yinying Z., Feng X., Yi X., Xuan F. & Haifeng G. (2020). The impact of artificial intelligence and blockchain on the accounting profession. IEEE Access 8(2020) pp.110461-110477

Zhang, Y., Xiong, F., Xie, Y., Fan, X., & Gu, H. (2020). The Impact of Artificial Intelligence and Blockchain on the Accounting Profession. IEEE Access, 8, 110461-110477.https://doi.org/10.1109/ACCESS.2020.3000505.

Zhou, Y., Li, T., Li, P., & Wen, S. (2023). Predicting Chinese students' destination choice intentions for international higher education in post-COVID-19: combining the image-value-behavior framework and protection motivation theory. Journal of Marketing for Higher Education, 1-24.