The Effect of Using Machine Learning Algorithms Alternatives on the Prediction Accuracy of Going Concern Opinion

Prepared by

Samar Fathy Abd El Gawad

Assistant Lecturer, Department of Accounting and Auditing

The Egyptian Institute of Alexandria Academy of Management and Accounting

Abstract:

The purpose of this paper is to analyze and examine the effect of machine learning algorithms alternatives on the prediction accuracy of going concern opinion and which one is more effective in predicting the accuracy of going concern opinion. **To achieve this purpose**, the research will address the accuracy of going concern opinion from a professional view, determinants of the accuracy of going concern opinion, measurements of accuracy of going concern opinion, machine learning from a professional view, and analysis the effect of the machine learning algorithms on the accuracy of going concern opinion.

In order to test the research hypotheses, the researcher will use the decision trees (DT), logistic regression, support vector machines (SVM). **The sample used in the current study consists of** 87 non-financial companies listed in Egyptian Stock Exchange during the period (2019-2021). **The research concludes that** SVM and Logistic regression has the highest accuracy to predict going concern doubts, where the accuracy rate is 86%, then the decision tree model doubts, where the accuracy rate is 79 %.

In light of the research objectives and its problem, and the results it concluded, the research recommends that auditors should be interested in developing their skills to be able to use artificial intelligence, such as machine learning, in issuing audit opinion, as they face some difficulties in using artificial intelligence in the audit field.
Regarding the proposed research areas, the most important of them are the following: (a) the effect of using data analytics on the prediction accuracy of going concern opinion, (b) The effect of artificial intelligence technologies on audit evidence, (c) The effect of machine learning on detecting misstatements on financial statements.

Keywords: Going concern prediction; Machine learning; Support vector machine (SVM); Decision tree; Logistic regression.

1-Introduction:

Accounting is a profession that adds value to the various stakeholders in the accounting firm. **The external audit is considered an interactive information system** that affects, and is also affected by, the variables of its operating environment and ends with the auditor's opinion on the financial statements.

In recent years many companies have been bankrupt, which led to losses to financial statement users, and thus the accuracy of the going-concern opinion of companies has received more attention from auditors. **Accuracy of going concern opinion** indicates that the auditor's opinion on going concern is based on a valid judgment in light of compliance with the relevant auditing standards and the Code of Ethics and professional conduct (Amr, 2022; Chi and Shen, 2022).

With regard to the auditor's professional judgment regarding going concern, the auditor evaluates the appropriateness of management's application of the going concern assumption since the accuracy of its opinion on going concern affects decision makers. Inaccuracy of the auditor's judgment regarding going concern lead to two types of error; type 1 error and type 2 error. First, **type 1 error (false rejection)** refers that the auditor issuing a qualified going concern opinion and the company did not go bankrupt in the subsequent year. **Second, type 2 error (false acceptance)** refers that the auditor issuing an unqualified going concern opinion, and the company went bankrupt in the subsequent year (Budisantoso et al., 2017).

Assessing going-concern doubts in companies is a complex process, which led to developing going-concern prediction models rather than traditional statistical methods. Some researchers have proposed constructing going-concern prediction models by data mining and machine-learning technologies, such as decision trees (DT), artificial neural networks (ANNs), and support vector machines (SVMs), which are the most commonly used to predict the going concern (Chi and Shen, 2022).

Therefore, the research problem can be expressed in how to answer the following question practically; what is the effect of machine learning algorithms alternatives (SVM, decision trees, and Logistic regression) on the prediction of going concern opinion, and which one is more accuracy in predicting of going concern opinion in non-financial companies listed in the Egyptian Stock Exchange.

The objective of this research is to analysis and examine the effect of decision tree, SVM and Logistic regression on the accuracy of going concern opinion prediction and which one is more effective in enhancing the accuracy of going concern opinion prediction in non-financial companies listed in the Egyptian Stock Exchange during the period (2019-2021).

The importance of the research is traced back to its alignment with the research that focused on improving the accuracy of going concern opinion. In addition, there is a significant scarcity of Egyptian research concerned with studying the machine learning algorithms alternatives and their impact on the accuracy of going concern opinion. **Despite the many research motives**, the most important of them is narrowing the research gap in this field, in addition to finding practical evidence on the existence of the relationship between the machine learning algorithms and the accuracy of going concern opinion in non-financial companies listed in the Egyptian Stock Exchange.

The limitation of this research, this research is limited to studying and testing the effect of the decision tree, SVM, and Logistic regression in predicting the accuracy of going concern opinion and which one is more effective in predicting the accuracy of going concern opinion in non-financial companies listed in the Egyptian Stock Exchange during the period (2019-2021). This research will focus on Decision trees, SVM, and Logistic regression as supervised learning, so other **supervised learning** methods like artificial neural networks (ANN), Genetic algorithm, and **Unsupervised learning methods** like Clustering, Unsupervised neural networks, and Association rules are beyond the scope of research.

Also, the other factors affecting the accuracy of going concern opinion, such as company size, leverage, growth rate, complexity, tenure, rotation, audit firm size, and experience are outside the extent of this research. In addition, the financial institutions and unlisted companies are outside the scope of this research. Finally, the generalizability of the results is conditioned by the criteria for selecting the research sample and the methodology used to test its hypotheses.

The remainder of this paper will be organized as follows: Section 2 discusses prior literature related to the accuracy of going concern opinion and machine learning algorithms, and the effect of the machine learning algorithms on the accuracy of going concern opinion and derives the hypotheses. Section 3 discusses the research methodology and design. Section 4 presents the empirical results and conclusion.

2- Literature Review and Hypotheses Development:

2-1 The Accuracy of Going Concern Opinion from Professional View:

Going concern assumption is one of the four basic accounting assumptions in the preparation of financial statements in accordance with generally accepted accounting assumptions (IAASB, 2015). IAS No. (1) explained that Conceptual Framework indicated that going concern assumption means that the company is a going concern and will continue in its operations for the prospective future. SAS No. (59) referred that the continuation of the company as a going concern is assumed in financial reporting in the absence of material information to the contrary. In the same context, some studies (Anggarini and Zulfikar, 2022; Chi and Shen, 2022;Hardies et al., 2018; Hafid et al., 2018) defined going concern as the assumption that the company will continue to operate for at least one year after the financial statement date.

Regarding the definition of Accuracy of going concern opinion, it indicates that the auditor’s opinion on going concern is based on a valid judgment in light of compliance with the relevant auditing standards and the Code of Ethics and professional conduct (Amr, 2022).

With regard to the auditor’s responsibilities, International Standard on Auditing (**ISA, No. 570**) states that the responsibilities of the auditor are to evaluate the appropriateness of management’s use of the going concern basis of accounting in the preparation of the financial statements. In addition to determine if there are some material events that lead to significant doubt on the company’s ability to continue as a going concern, and determine whether management has already performed a preliminary assessment of the company’s ability to continue as a going concern and (a) If the management performed such an assessment, the auditor should discuss the assessment with management and determine whether management has identified any material event and, if so, management’s plans to address them; or (b) If the management did not perform such an assessment, the auditor shall discuss with management the basis for the intended use of the going concern basis of accounting. In addition, the auditor should issue his judgment on the sufficient of the audit procedures it performs, and whether there is a need for additional audit procedures, in the event that management delays the approval of financial statements due to going concern issues. Finally, Issuing audit opinion.

Concerning the American standard (AS No. 2415), it is consistent with (ISA, No. 570) in determining the auditor's responsibilities in the audit of financial statements relating to going concern, **but it differs in some aspects**, in which the auditor is not required to perform additional audit procedures if the management's delay in approving the financial statements due to reasons related to the company's ability to continue as a going concern. Also, **Egyptian Accounting Standard No 570 agrees with (ISA, No. 570)** concerning the auditor responsibilities in the audit of financial statements relating to going concern.

The researcher concluded that the accuracy of the going concern opinion refers to the auditor's compliance with his responsibilities in accordance with auditing standards, which reflects positively on the accuracy of his judgment.

2-2 Determinants of the Accuracy of Going Concern Opinion:

Numerous studies (Anggarini and Zulfikar, 2022; Hardies, 2020; Simamora and Hendarjatno, 2019; Hafid et al., 2018; Hapsoro and Santoso, 2018; Gallizo and Saladrigues, 2016; Haron et al., 2009) indicated that there are many factors affecting the accuracy of Going Concern Opinion, which can be divided into three categories. **First, the characteristics of the audit client company**, the most important of which are; company size, liquidity, leverage, profitability, cash flow of operation, growth rate, financial position, governance commitment, complexity, and quality of financial reports. **Second, the qualitative characteristics of external auditors**, the most important of which are; audit firm size, experience, tenure, rotation, reputation, audit fees, independence, and litigation risk. **Finally, machine learning techniques** such as Neural networks (NN), Support vector machine (SVM), and Decision trees (DT).

In the present study, we will focus on the factors related to the machine learning techniques and their effect on the accuracy of going concern opinion.

2-3 Measurements of the Accuracy of Going Concern Opinion:

According to (Amr, 2022; Budisantoso et al., 2017; Gallizo and Saladrigues, 2016; Geiger & Rama, 2006), **accuracy of going concern opinion is measured by** a dummy variable equal to (1) if the auditor issued correct opinion about the going concern (true acceptance and rejection) and (0) otherwise (false acceptance and rejection), this is by the extent of its agreement or disagreement with bankruptcy prediction models.

2-3-1 Traditional Altman Z-score:

The Traditional Altman Z-Score model was developed by Altman in 1968. This model helps to predict bankruptcy in addition to detecting earning manipulation. Traditional Altman Z-score consists of five variables that can be calculated with the information provided in the financial statements of the company (Parikh and Shah, 2022; Somayyeh, 2015). Traditional Altman Z-score is calculated from the following formula:

$$\mathbf{Z=1.2X1+1.4\ X2+3.3\ X3+0.6\ X4+1\ X5}$$

where: X1 = Working capital / Total assets, X2 = Retained earnings / Total assets, X3 = Earnings before interest and tax / Total assets, X4= Market Value of Equity / Total liabilities L, X5= Net Sales/Total assets.

If Z is greater than 2.67, it indicates that the possibility of not being bankrupt is very high (safe zone), and If Z is less than or equal to 1.81, it indicates that the possibility of being bankrupt is very high (distress zone), but if Z- Score is greater than 1.81 and less than 2.67, it indicates that the company in (grey zone) (Parikh and Shah, 2022; Somayyeh, 2015).

2-3-2 Zmijewski (1984) X-score:

The Zmijewski (1984) X-score model is the most commonly used model by researchers (Grice and Dugan, 2003). X-score consists of three variables that can be calculated with the information provided in the financial statements of the company (Puspaningsih & Analia, 2020; AlAli et al., 2018). Zmijewski is calculated from the following formula:

$$\textbf{Zmijewski X-score} = -4.3 - 4.5X1 + 5.7X2 + 0.004X3$$

Where X1 (ROA)= Net income / Total assets, X2 = Total liabilities / Total Assets, X3(current ratio) = Current assets / Current liabilities

If the X-score is negative (X-Score>0), it indicates that the possibility of not being bankrupt is very high (healthy condition), and if the x-score is positive (X-Score≥ 0), it indicates that the possibility of being bankrupt is very high (distress zone) (Puspaningsih & Analia, 2020; AlAli et al., 2018).

2-3-3 Ohlson O-score:

The Ohlson O-score was developed by James Ohlson in 1980 as an alternative to the Altman Z-score for predicting financial distress (Bansal, 2020). O-score consists of nine variables that can be calculated with the information provided in the financial statements of the company. O-score is calculated from the following formula:

O = -1.32 - .407X1 + 6.03X2 - 1.43X3 + .0757X4 - 2.37X5 - 1.83X6 + 0.285X7 - 1.72X8 - .521X9

Where: X1 = ln(total assets/GNP price-level index), X2 = total liabilities/total assets, X3 = working capital/total assets, X4 = current liabilities/current assets, X5 = one if total liabilities exceed total assets, else zero, X6 = net income/total assets, X7 = funds provided by operations/total liabilities, X8 = one if net income was negative for the last two years, else zero, X9 = $(NI_t - NI_{t-1})/(|NI_t| + |NI_{t-1}|)$, where N_{it} is net income (Bansal, 2020).

2-3-4 Beneish M-Score:

The Beneish M-Score model was developed by Messod Beneish in 1999. This model can classify the companies as fraudulent and non-fraudulent companies. M-Score consists of eight variables that can be calculated with the information provided in the financial statements of the company (Parikh and Shah, 2022; Somayyeh, 2015; Beneish, 1999). M-Score is calculated from the following formula:

M = −4.84 + 0.92DSRI + 0.528GMI + 0.404AQI + 0.892 SGI + 0.115 DEPI − 0.172 SGAI + 4.679 TATA − 0.327 LVGI

Where: **DSRI**= $(\mathit{Net\ Receivables}_t / \mathit{Sales}_t) / (\mathit{Net\ Receivables}_{t-1} / \mathit{Sales}_{t-1})$, **GMI**= $[(\mathit{Sales}_{t-1} - \mathit{Cost\ of\ Goods\ sold}_{t-1}) / \mathit{Sales}_{t-1}] / [(\mathit{Sales}_t - \mathit{Cost\ of\ Goods\ sold}_t) / \mathit{Sales}_t]$, **AQI=** $[1 - (\mathit{Current\ Assets}_t + \mathit{PPE}_t) / \mathit{Total\ Assets}_t)] / [1 - (\mathit{Current\ Assets}_{t-1} + \mathit{PPE}_{t-1} / \mathit{Total\ Assets}_{t-1})]$, **SGI=** $\mathit{Sales}_t / \mathit{Sales}_{t-1}$, **DEPI=** $[\mathit{Depreciation}_{t-1} / (\mathit{Depreciation}_{t-1} + \mathit{PPE}_{t-1})] / [\mathit{Depreciation}_t / (\mathit{Depreciation}_t + \mathit{PPE}_t)]$, **SGAI= [**$\mathit{sales,\ general\ and\ administrative\ expenses}_t / \mathit{sales}_t$] / [$\mathit{Sales,\ general\ and\ administrative}$

$expenses_{t-1}/Sales_{t-1}$], **TATA=** $Total\ Accruals_t\ /\ Total\ Assets_t$, **LVGI=** [$LTDt + Current\ liabilities_t/Totals\ Assets_t$] / [$LTD_{t-1} + Current\ liabilities_{t-1}/\ Total\ Assets_{t-1}$].

If M-Score value is greater than -2.22, it indicates that the company is likely to be manipulating its financial statements, and if the M-Score value is less than -2.22, it indicates that the company does not manipulate its financial statements (Parikh and Shah, 2022; Somayyeh, 2015).

2-3-5 Altman for Emerging Economies (2005) Model:

Altman for emerging economies developed in the mid-1990s. It differs from traditional Altman model, that it does not require companies to be listed on the stock exchange, as it replaces the market value of equity with the book value of equity (x4). In addition, it can be applied to nonmanufacturing companies, manufacturing, and is relevant for private and public firms in developing countries. z-score is calculated from the following formula:

Z= 6.56 X1 + 3.2 X2 +6.72 X3 +1.05 X4 + 3.25
where: X1 = Working capital / Total assets , X2 = Retained earnings / Total assets, X3=Operating income / Total assets, X4= Book value of equity /Total liabilities.
If Z is greater than 2.6, it indicates that the possibility of not being bankrupt is very high (safe zone), if Z is between 1.1 and 2.6 is the (grey area) and if Z is smaller than 1.1 means that the possibility of being bankrupt is very high (distress zone) (Amr,2022; Altman,2005).

2-4 Machine Learning from a Professional View:

Machine learning is a subcategory of artificial intelligence[1]. It can be interpreted as algorithms that identify and extract patterns from the provided data (Kelleher & Tierney, 2018). Khanzode and Sarode (2020) indicated that machine learning is a scientific study of algorithms and statistical models when the computer system used to perform a specific task. Machine learning has generated many success stories over the past few years. For example, machine learning algorithms are used to identify fraudulent credit card transactions, and customer segmentation in targeted advertising campaigns and in medicine to detect tumors (Gierbl, 2021).

[1] Artificial intelligence (AI) is an integration of computer science and physiology, and it concerned with making computers behave like humans more human (Khanzode and Sarode, 2020).

When performing data analytics with machine learning algorithms, the database will be divided into training, validation, and testing sets (Kelleher & Tierney, 2018). The training set is used to train the models, and then the validation set is used to compare the performance of several models on new data not used for training, to find the best one. Once the most appropriate model has been chosen, the validation and training sets are combined to train the model on a larger dataset. Finally, the testing set is then used to evaluate the performance of the final model (Gierbl, 2021).

Machine learning algorithms can be divided into two categories: supervised learning and unsupervised learning. For **supervised methods**, the datasets include “labeled” examples with target information (e.g., fraud, non-fraud, bankrupt, non-bankrupt). There are a lot of **supervised methods** such as the Nearest neighbor, Naïve Bayes, Bayesian belief networks, Decision trees and random forests, Artificial neural networks / deep learning, Support vector machines, and Linear and logistic regression. For **unsupervised learning methods**, the datasets used are without labeled target information. There are a lot of **unsupervised methods**, such as Clustering, Unsupervised neural networks, and Association rules (Gierbl, 2021).

With respect to using machine learning techniques in going concern prediction, previous studies (Chi and Shen, 2022; Jan, 2021; Goo et al., 2016) have predicted going concern by using machine learning logarithms such as artificial neural network (ANN), decision trees (DT), and support vector machine (SVM).

In the current research, the researcher will depend on decision trees (DT), Support vector machines, and logistic regression to predict going concern doubts because the Supervised learning algorithms have been very effective in performing prediction tasks like fraud detection and bankruptcy prediction (Zhang, 2018).

2-5 Analysis of the Effect of the Machine Learning Algorithms Alternatives on the Accuracy of Going Concern Opinion, and Research Hypotheses Development:

Machine learning is one way to achieve Artificial Intelligence (AI), which is the science of making computers do things that require intelligence when done by humans (Alpaydin, 2020). Machine learning is the programming of computers to automate the discovery of patterns from existing data or past experiences that might be difficult to find otherwise

(Alpaydin, 2020; Cecchini et al. 2010). Also, it is used to solve problems for which humans do not know an explicit answer, such as recognizing patterns in images or classifying spam emails from legitimate ones (Alpaydin, 2020).

Several previous studies (Chi and Shen 2022; Jan 2021; Budisantoso et al., 2017; Goo et al., 2016; Gallizo and Saladrigues, 2016; Geiger & Rama, 2006) have indicated that there are several techniques that can auditor used to predict the going concern opinion. These methods are divided into traditional methods and machine learning methods.

With regard to the traditional methods that affect the accuracy of the auditor's opinion on going concern, some studies (Budisantoso et al., 2017; Gallizo and Saladrigues, 2016; Geiger & Rama, 2006) have indicated that many **bankruptcy prediction models** affect the accuracy of the auditor's opinion on going concern, such as Traditional Altman Z-Score model, Pustylnick P-Score, Beneish M-Score, Zmijewski (1984)X-score, Ohlson O-score, and altman for emerging economies model.

While on the other side, there are some studies (Chi and Shen 2022; Jan 2021; Goo et al., 2016**) have indicated that there are machine learning methods that** improve the predictions of the auditor's going concern opinion, which lead to increase in the accuracy of going concern opinion, such as artificial neural networks (ANN), decision trees (DT), and support vector machine (SVM) which are supervised methods.

Perols (2011) compares the performance of six machine learning models in detecting financial statement fraud, and the results show that logistic regression and support vector machines perform well relative to an artificial neural network, bagging, C4.5, and stacking. **In the same context**, Jan (2021) constructs going concern prediction models to help auditors to make more accurate judgments on going concern opinion decisions by two commonly used deep learning algorithms; recurrent neural network (RNN), and deep neural networks (DNN).

Chi and Shen (2022) are consistent with Jan (2021), where Chi and Shen (2022) referred that traditional models have disadvantages and high error rates in giving going-concern opinions, so it is necessary to use effective and accurate models to predict going-concern opinion. Machine learning technology is one of the artificial intelligence tools used to overcome the disadvantages of these traditional models and reduce judgment errors. This research constructs the going-concern prediction models by three machine-learning algorithms, namely extreme gradient boosting (XGB), ANN, and SVM. The prediction accuracy rate of all models is higher than 80%. This research is based on the financial and non-financial variables to predict going concern doubts, such as (debt ratio, return on stockholders' equity, return on total assets, earnings per share, total liabilities/stockholders' equity . . . etc).

On the other hand, previous studies indicated that although the efficacies of AI are significant, there are some difficulties in using it by auditors, as artificial intelligence tools are easier to be used by programmers. So, the auditors need to acquire more skills to be able to use AI. Also, it is often hard to understand the nature of the problem and the solution. In addition, sometimes, it can be misused, which results in mass destruction (Khanzode and Sarode, 2020; Chowdhury and Sadek, 2012).

It is clear that there is a paucity, within the limits of the researcher's knowledge, in the studies that examined the effect of machine learning algorithms on the accuracy of going concern opinion and which one algorithm is more accuracy, which represents a motivation to test this relationship in the present study and derive the hypothesis of the research as follows:

H1: Decision tree as one of the machine learning algorithms is more effective than SVM and Logistic regression in enhancing the accuracy of going concern opinion prediction for companies listed on the Egyptian Stock Exchange.

3- Research Methodology:

To achieve the objective of the research and then test its hypothesis, the researcher will depend on an empirical study. The researcher will present the following: The **objectives** of the empirical study, the **population and sample** of the research, the **research model**, description and measurement of variables, research tools and procedures, statistical analysis tools. This is as follows:

The empirical **study aims to test research hypothesis in the Egyptian business and professional practice environments, and to find practical evidence of the validity of the relationship under study.**

3-1 Empirical Study Objectives:

The empirical study aims to test research hypothesis in Egyptian business and professional practice environments to find practical evidence of the validity of the relationship under study.

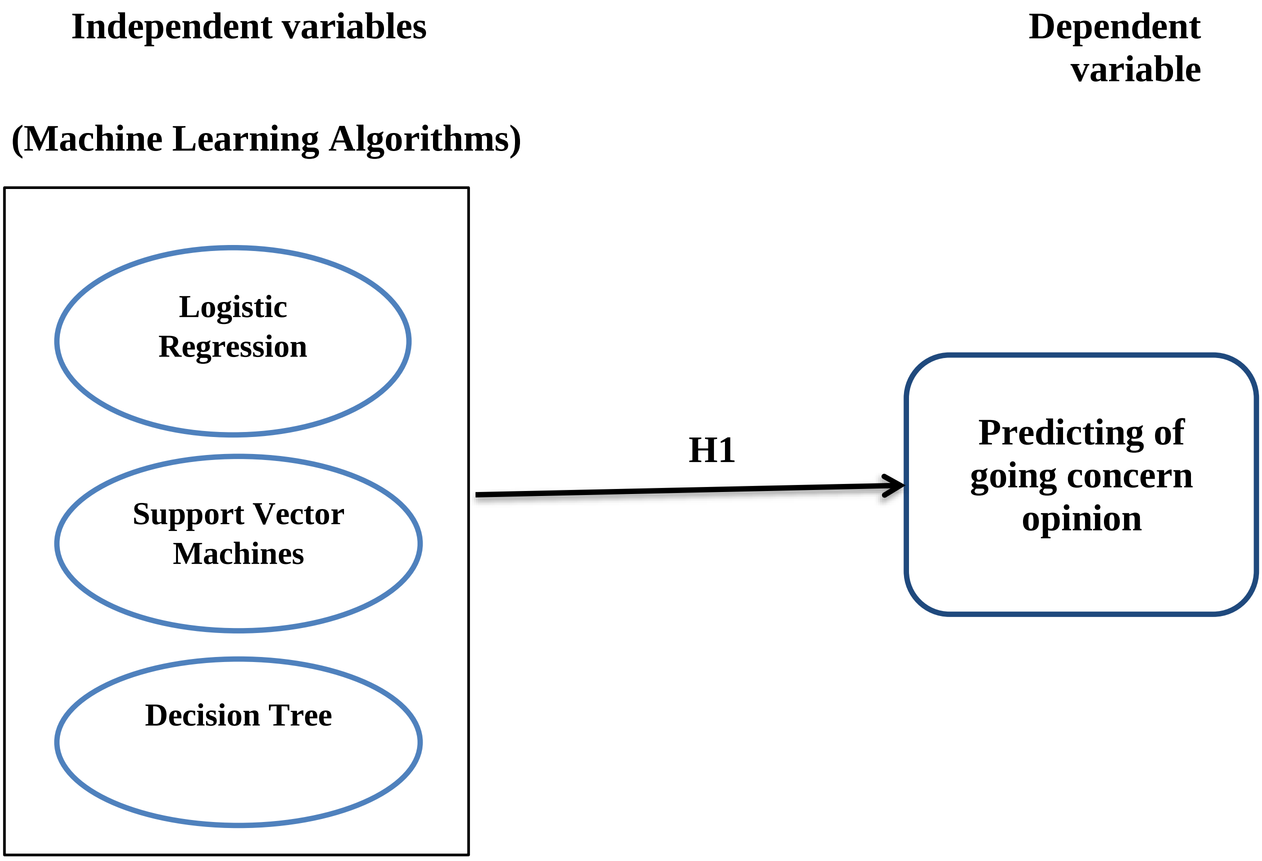
3-2 Population and Sample Selection:

The population of the study consists of non-financial firms listed in the Egyptian Stock Exchange during the period from 2019 to 2021. The sample consists of 87 firms. The number of observations is 254 firm-year observations, where the researcher follows the Firm-Year-Observation approach according to (Haw, et al., 2014). These observations were divided into 54 observations related to going concern opinion and 200 related to non-going concern opinion. The sample excluded the financial institutions because the nature of their activities are different from the nonfinancial institutions, the separate laws they follow (Perols 2011), also excluded firms with incomplete financial reports in addition to firms whose financial reports were prepared in foreign currency.

3-3 Research Model &Measurement of Variables:

3-3-1 Research Model:

The hypothesis of the research show that the independent variables **are three machine learning algorithms (Logistic Regression, Support Vector Machines, Decision Tree),** while the dependent variable is **predicting the accuracy of going concern opinion**. According to these variables, the research model is as following:



3-3-2 Measurement of Variables:

3-3-2-1 Dependent Variable:

Going concern opinion can be defined as “the auditor’s opinion on going concern is based on a valid judgment in light of compliance with the relevant auditing standards and the Code of Ethics and professional conduct”(Amr, 2022). It is measured by a dummy variable equal to (1) if the firm received the audit opinion of going concern doubt and (0) if the firm did not receive the audit opinion of going concern doubt.

3-3-2-2 Independent variables (Machine learning models to predict going concern):

Machine learning is a scientific study of algorithms and statistical models when the computer system used to perform a specific task (Khanzode and Sarode, 2020). There are a lot of supervised machine learning models to predict going concern such as support vector machines, Neural networks, logistic regression, and Decision tree. The researcher will depend on decision trees (DT), logistic regression, support vector machines to predict going concern doubts (Chi and Shen, 2022). The model consists of 19 variables, including 16 financial variables and 3 non-financial variables, which are summarized in Table 1 (Chi & Shen, 2022; Jan, 2021).

Table 1: Research variables

Variable classification	Code	Variable Name	Variable Definition or Calculation Equation
Financial variable	X1	Debt ratio	Total liabilities / Total assets
Financial variable	X2	Quick ratio	Quick assets / Current liabilities
Financial variable	X3	Current ratio	Current assets / Current liabilities
Financial variable	X4	D/E ratio	Total liabilities /Total equity
Financial variable	X5	Current liabilities ratio	Current liabilities / Total liabilities
Financial variable	X6	Ratio of current assets to total liabilities	Current assets / Total liabilities
Financial variable	X7	Ratio of long-term funds to fixed assets	(Stockholders' equity + long-term liabilities/ fixed assets
Financial variable	X8	Interest coverage ratio	EBIT / Interest expense
Financial variable	X9	ROA	[Net income + interest expense* (1 - tax rate)] / Average total assets
Financial variable	X10	ROE	Net income / Average total equity
Variable classification	**Code**	**Variable Name**	**Variable Definition or Calculation Equation**
Financial variable	X11	Total assets turnover	Net Sales / Total assets
Financial variable	X12	Accounts receivable turnover	Net sales / Average accounts receivable
Financial variable	X13	Inventory turnover	Cost of goods sold / Average inventory
Financial variable	X14	EPS	Net income / Shares of common stock
Financial variable	X15	Gross margin	Gross profit / Net sales
Financial variable	X16	Operating income ratio	Operating income / Net sales
Non-Financial variable	X17	Stockholding ratio of major shareholders	Stockholding ratio of major shareholders / Shares of common stock
Non-Financial variable	X18	Pledge ratio of directors and supervisors	Pledge ratio of directors and supervisors / Shares of common stock
Non-Financial variable	X19	Audited by BIG4 (the big four CPA firms) or not	1 for companies audited by BIG4, otherwise is 0

3-3-3 Data Collection and Analysis:

In relation to data collection, this research depends on the secondary data that collected from the financial reports for the firms that used in this study, where the financial reports include the financial statements. Financial and non-financial ratios used in the study were calculated, then emptying this data in a Microsoft excel sheet and analyzing it by using the **Anaconda program using python**. To test the research hypothesis, the researcher will use the decision trees (DT), logistic regression, support vector machines, where its used to predict discrete variables and it continuously partitions data into two subsets by the binary method, where the dependent variable going concern opinion will be measured using binary variables (Jan, 2021).

Regarding data analysis, when using machine learning to analyze the data, the dataset should be split into three groups: training, validation, and testing (Jan, 2021). The training set is used to train the models. The validation set is a set of data, separate from the training set, that is used to validate our model performance during training and find the best one. Once the most appropriate model has been selected, the validation and training sets are combined to train the model on a larger dataset. The testing dataset is then used to evaluate the performance of the final model (Gierbl, 2021). In the modeling process, 70% of data is used for modeling, where 75% of this dataset is used as a training dataset (70% * 75%= 52.5%) and 25% is used as a validation dataset (70% *25% = 17.5%), while the remaining 30% of the data are used as testing datasets to test the model (Jan, 2021).

We will apply three alternatives of Machine Learning Algorithms (Decision tree, logistic regression, and support vector machines) to determine their effect on the prediction of Going Concern Opinion, in addition to evaluating the performance of each model by using accuracy rate, the error rate in addition to confusion matrix indicators which are accuracy, precision, recall, and F1-score.

3-3-4 Statistical Methods and Research Models:

The decision trees (DT), logistic regression, support vector machines were used to test the hypothesis of the research as follows:

3-3-4-1 Testing the Hypothesis:

The objective of this hypothesis is to test whether decision tree as one of machine learning algorithms are more effective than SVM and Logistic regression in enhancing the accuracy of going concern opinion prediction according to the following equations:

GCO = β0 + β1 ML (Logistic regression) + £it (1)

GCO = β0 + β1 ML (SVM) + £it (2)

GCO = β0 + β1 ML (DT) + £it (3)

Where:

GCO: Auditor's going concern opinion

ML: Machine learning

DT: Decision tree

SVM: support vector machines

£ = error term, it = firm i in year t.

4- Empirical Findings:

4-1 Descriptive Statistics of Data:

The descriptive statistics of the dependent variable, financial and nonfinancial ratios are shown in table (2). It is clear from the table that the mean value of the debt ratio, Current liabilities ratio and Total assets turnover (.42; .83; .58) respectively, is higher than its standard deviation (.30; .20; .55), respectively, and that the minimum value and the maximum value fluctuates between (0.00- 3.18), (.06-1.00), (-.03- 3.21), respectively. **And in contrast** the mean value of Quick ratio, Current ratio, D/E ratio, Ratio of current assets to total liabilities, Ratio of long-term funds to fixed assets, Interest coverage ratio, ROA, ROE, Accounts receivable turnover, Inventory turnover, EPS, Gross margin, Operating income ratio, Stockholding ratio of major shareholders, and Audited by an audit firm partner in BIG4 or not (2.46; 3.55; 0.92; 2.81; 101.72; .63; -.2583; -23.36; .02; 7.08; 9.09; 1.29; 0.19; -6.20; 4.16; .25) respectively, are less than their standard deviation (7.66; .8.62; 1.11; 7.87; 11112.28; 1089.78; 1356.51; 0.48; 12.35; 20.75; 16.89; 0.43; 95.82; 14.49; .44) respectively, and their minimum and maximum values range between (.00- 107.35), (.00- 107.35), (-3.14- 8.42), (.00- 107.35), (-127.5- 17646.07), (-14500.25- 4919.37), (-18646.49- 4679.35), (-6.79- 0.70), (-1.19, 101.95), (0.00- 218.99), (-166.90- 184.97), (-4.6- 1.85), (-1527.04, 3.18), (0.00, 80), (0.00, 1), respectively.

Table 2: Descriptive Statistics

Variables	Mean	SD	Min	Max	N
GC	.2	.4	0.0	1.0	254
Debt ratio	.42	.30	.00	3.18	254
Quick ratio	2.46	7.66	.00	107.35	254
Current ratio	3.55	8.62	.00	107.35	254
D/E ratio	.92	1.11	-3.14	8.42	254
Current liabilities ratio	.83	.20	.06	1.00	254
Ratio of current assets to total liabilities	2.81	7.87	.00	107.35	254
Ratio of long-term funds to fixed assets	101.72	11112.28	-127.5	17646.07	254
Interest coverage ratio	.63	1089.78	-14500.25	4919.37	254
ROA	-23.36	1356.51	-18646.49	4679.35	254
ROE	.02	0.48	-6.79	0.70	254
Total assets turnover	0.58	0.55	-.03	3.21	254
Accounts receivable turnover	7.08	12.35	-1.19	101.95	254
Inventory turnover	9.09	20.75	0.00	218.99	254
EPS	1.29	16.89	-166.90	184.97	254
Gross margin	0.19	0.43	-4.6	1.85	254
Operating income ratio	-6.20	95.82	-1527.04	3.18	254
Stockholding ratio of major shareholders	4.16	14.49	0.00	80	254
Pledge ratio of directors and supervisors	.29	.29	0.00	1.37	254
Audited by an audit firm partner in BIG4 (the big four CPA firms) or not	.25	.44	0.00	1	254

4-2 Hypothesis Testing:

To compare the prediction performance among the models, this research uses the accuracy rate, and GC sample prediction error rate. The confusion matrix is also used in this research. The indicators of the confusion matrix are accuracy, precision, recall, and F1-score to evaluate the performance of a model.

Table 3: Summary of Going Concern Prediction models

Model	**Accuracy Rate**	**Error Rate**	**Precision Rate**	**Recall Rate**	**F1 Score Rate**	**Type 1 error**	**Type 2 error**
SVM	86%	14%	86%	100%	92%	0 observation	11 observation
Logistic regression	86%	14%	86%	100%	92%	0 observation	11 observation
DT	7٩%	21%	86%	91%	88%	6 observation	10 observation

Table (3) reports the results of **Prediction Going Concern models** as follows: The results of the models above show that **SVM and Logistic regression have the highest accuracy to predict going concern doubts,** where **the accuracy rate**[1] of the test dataset of SVM and Logistic regression (86%) is **higher than** the accuracy rate of the Decision tree (79%). While the **error rate**[2] of SVM and Logistic regression (14%) is **less than** Decision tree (21%). Also, the precision rate[3], recall rate[4], and F1 score rate[5] of **SVM and Logistic regression** (86%, 100%, 92%), respectively, are **higher than** the Decision tree which is (86%, 91%, 88%), respectively.

---

[1] **The Accuracy rate** is calculated by the following equation: the number of correct predictions / total number of predictions (Chi and Shen, 2022).

[2] **The error rate** is calculated by the following equation: the number of incorrect predictions / total number of predictions (Chi and Shen, 2022).

[3] **Precision** measures how many of the positive predictions made are correct (true positives). It is calculated by the following equation: true positive / total number of positive predicted (true positive predictions+ false positive predictions) (Chi and Shen, 2022).

[4] **Recall** measures how many positive cases the classifier correctly predicted, over all the positive cases in the data. It is calculated by the following equation: recall = true positive /(true positive actual + false actual negative actual) (Chi and Shen, 2022).

[5] **F1-score** is the harmonic mean (average) of the precision and recall. It is calculated by the following equation: F1-score = 2* precision * recall /(precision + recall) (Chi and Shen, 2022).

Also, **the SVM and logistic regression models** show that zero non-going concern financial statement is incorrectly classified as going concern (false positive), so the type 1 error is zero. In addition, 11 going concern financial statements are incorrectly classified as non-going concern financial statements (false negative), so type 2 error is 11 financial statements.

In the same context, **the Decision tree model**, 6 non-going concern financial statement is incorrectly classified in going concern (false positive), so the type 1 error is 6. In addition, 10 going concern financial statements are incorrectly classified as non-going concern financial statements (false negative), so type 2 error is 10 financial statements.

According to the above results, we will reject H1, where SVM and Logistic regression is more effective than Decision tree in predicting the accuracy of going concern opinion**.** The result of this study disagrees with the result concluded by (Chen et al., 2014) which found that decision tree model is more accuracy in predicting of going concern doubts.

5- Discussion and Conclusion:

External Auditors audit companies' financial statements and issue their opinions. These audit opinions are very important for stakeholders, and financial markets, especially investors. So, it is necessary to establish more accurate going-concern doubt prediction models.

This study uses three machine learning algorithms; SVM, Logistic regression, and decision tree to predict going concern doubts**.** When using machine learning to analyze the data, the dataset should be split into three groups: training, validation, and testing. The training set is used to train the models. The validation set is a set of data, separate from the training set, that is used to validate our model performance during training and find the best one. Once the most appropriate model has been selected, the validation and training sets are combined to train the model on a larger dataset. The testing dataset is then used to evaluate the performance of the final model. Financial and non-financial ratios used in the study were calculated, and analyzing it by using the **Anaconda program using Python**.

The result of the present study showed **that SVM and Logistic regression has the highest accuracy to predict going concern doubts**, where the accuracy rate, precision rate, recall rate, and F1 score rate are

(86%, 86%, 100%, 92%), respectively, **then the decision tree model doubts, where the accuracy rate,** precision rate, recall rate, and F1 score rate are (79 %; 86%, 91%, 88%), respectively.

6- Research Recommendations and Future Research Opportunities:

In light of the research objectives and its problem, and the results it concluded, the research recommends that auditors should be interested in developing their skills to be able to use artificial intelligence, such as machine learning, in issuing audit opinion, as they face some difficulties in using artificial intelligence in the audit field.

Regarding the proposed research areas, the most important of them are the following: (a) the effect of using data analytics on the prediction accuracy of going concern opinion, (b) The effect of artificial intelligence technologies on audit evidence, (c) The effect of machine learning on detecting misstatements on financial statements.

References:

AlAli, M.S., AlShamali, M.M., AlAwadhi, K.M. and AlSabah, A.M., 2018. The use of Zmijewski model in examining the financial soundness of oil and gas companies listed at Kuwait Stock Exchange. ***International Journal of Economics, Commerce and Management Research Studies***, *1*(2), pp.15-21.

Alpaydin, E., 2020. ***Introduction to machine learning***. MIT press.

Altman, E.I., 2005. An emerging market credit scoring system for corporate bonds. ***Emerging markets review***, *6*(4), pp.311-323.

Amr, A.M., 2022**. The effect of financial and non-financial aspects of both audit firm and its client on the auditor`s professional judgement accuracy regarding going concern- An applied study on listed companies in EGX** (Doctoral dissertation, Alexandria University).

Anggarini, D. and Zulfikar, Z., 2022. Factors Affecting Audit Opinion Going Concern. In ***International Conference on Economics and Business Studies (ICOEBS 2022)***, pp. 8-14.

Bansal, R., Kashyap, S., Pranav, P., Krishna, U., Ahmad, Z. and Chauhan, A., 2020. Altman and Ohlson model in predicting distress of Indian companies: A comparison of models. ***European Journal of Molecular and Clinical Medicine***, *7*, pp.4158-4167.

Beneish, M. D., 1999. The detection of earnings manipulation. ***Financial Analysts Journal***, *55*(5), pp.24-36.

Budisantoso, T., Rahmawati, R., Bandi, B. and Probohudono, A.N., 2017. Audit opinion accuracy, corporate governance and downward auditor switching: A study of association of Southeast Asian nations economics community. ***International Journal of Economics and Financial Issues***, *7*(5), pp.530-540.

Byusi , H. and Achyani , F.,2018. Determinants of Audit Going Concern Opinion (Empirical Study of Real Estate and Property Companies Listed on the Idx In 2013-2015), ***Indonesian Accounting and Finance Research***, *3* (1), pp. 13–28.

Cecchini, M., Aytug, H., Koehler, G.J. and Pathak, P., 2010. Detecting management fraud in public companies. ***Management Science***, *56*(7), pp.1146-1160.

Chen, S., Goo, Y.J.J. and Shen, Z.D., 2014. A hybrid approach of stepwise regression, logistic regression, support vector machine, and decision tree for forecasting fraudulent financial statements. ***The Scientific World Journal***, pp.1-9.

Chi, D.J. and Shen, Z.D., 2022. Using hybrid artificial intelligence and machine learning technologies for sustainability in going-concern prediction. ***Sustainability***, *14*(3), pp.1-18.

Chowdhury, M. and Sadek, A.W., 2012. Advantages and limitations of artificial intelligence. ***Artificial intelligence applications to critical transportation issues***, *6*(3), pp.360-375.

Gallizo, J.L. and Saladrigues, R., 2016. An analysis of determinants of going concern audit opinion: Evidence from Spain stock exchange. ***Intangible Capital***, *12*(1), pp.1-16.

Geiger, M.A. and Rama, D.V., 2006. Audit firm size and going-concern reporting accuracy. ***Accounting horizons***, *20*(1), pp.1-17.

Gierbl, A.S., 2021. **Data analytics in external auditing** (Doctoral dissertation, Universität St. Gallen).

Goo, Y.J.J., Chi, D.J. and Shen, Z.D., 2016. Improving the prediction of going concern of Taiwanese listed companies using a hybrid of LASSO with data mining techniques. ***SpringerPlus***, *5*(1), pp.1-18.

Grice Jr, J.S. and Dugan, M.T., 2003. Re-estimations of the Zmijewski and Ohlson bankruptcy prediction models. ***Advances in Accounting***, *20*, pp.77-93.

Hapsoro, D. and Santoso, T.R., 2018. Does audit quality mediate the effect of auditor tenure, abnormal audit fee and auditor's reputation on giving going concern opinion?. ***International Journal of Economics and Financial Issues***, *8*(1), pp.143-152.

Hardies, K., 2020. Auditor independence impairment: bonding between clients and individual engagement partners. ***Spanish Journal of Finance and Accounting***, *49*(4), pp.427-451.

Hardies, K., Vandenhaute, M.L. and Breesch, D., 2018. An analysis of auditors' going-concern reporting accuracy in private firms. ***Accounting Horizons***, *32*(4), pp.117-132.

Haron, H., Hartadi, B., Ansari, M.A. and Ismail, I., 2009. Factors influencing auditors' going concern opinion. ***Asian Academy of Management Journal***, 14(1), pp.1–19.

Haw, I.M., Lee, J.J. and Lee, W.J., 2014. Debt Financing and Accounting Conservatism in Private Firms. ***Contemporary Accounting Research***, 31(4), pp.1220-1259.

International Accounting Standards Board (IASB), 2014. **Presentation of Financial Statements**. International Accounting Standard (IAS) 1. Available at : www.iasb.org.

International Auditing and Assurance Standard Board (IAASB), 2015**. Going Concern**. International standards on auditing (ISA) 570 revised. Available at : www.ifac.org.

Jan, C.L., 2021. Using deep learning algorithms for CPAs’ going concern prediction. ***Information***, *12*(2), p.73.

Kelleher, J.D. and Tierney, B., 2018. Audit Considerations in Respect of Going Concern in The Current Economic Environment. ***Data science***. MIT Press.

Khanzode, K.C.A. and Sarode, R.D., 2020. Advantages and disadvantages of artificial intelligence and machine learning: A literature review. ***International Journal of Library & Information Science (IJLIS)***, *9*(1), pp.30-36.

Parikh, M.Y. and Shah, K., 2022. Financial Statement Fraud Detection Models: An Exploratory Study. ***A Global Journal of Social Sciences***, 5(3), pp. 21-26.

Perols, J., 2011. Financial statement fraud detection: An analysis of statistical and machine learning algorithms. ***Auditing: A Journal of Practice & Theory***, *30*(2), pp.19-50.

Public Company Accounting Oversight Board (PCAOB). (2015). **Considerations of Entity`s Ability to Continue as a Going concern**. Auditing Standard (AS) 2415 .Available ***at www. pcaobus.org.***

Puspaningsih, A., and A. P.Analia, 2020. The Effect of Debt Default, Opinion Shopping, Audit Tenure and Company's Financial Conditions on Going-concern Audit Opinions. ***Review of Integrative Business and Economics Research***, *9*(2), pp.115-127.

Simamora, R.A. and Hendarjatno, H., 2019. The effects of audit client tenure, audit lag, opinion shopping, liquidity ratio, and leverage to the going concern audit opinion. ***Asian Journal of Accounting Research***, *4*(1), pp.145-156.

Somayyeh, H.N., 2015. Financial ratios between fraudulent and non-fraudulent firms: Evidence from Tehran Stock Exchange. ***Journal of Accounting and Taxation***, *7*(3), pp.38-44.

Zhang, C.A., 2018. Predict Audit Quality Using Machine Learning Algorithms. ***Available at SSRN***, https://ssrn.com/abstract=3449848.