Leveraging Artificial Intelligence for Automated Transaction Mapping to Accounting Standards

Linda William, Renyu Lu, Ester Goh and Woelly William

September 26, 2024

Leveraging Artificial Intelligence for Automated Transaction Mapping to Accounting Standards

Linda William[1], Renyu Lu[1], Ester Goh[1], and Woelly William[2]
[1] School of Informatics & IT, Temasek Polytechnic, Singapore
[2] Edugame Analytics Pte Ltd., Singapore
lwilliam@tp.edu.sg, 2202560E@student.tp.edu.sg, egohgh@tp.edu.sg, woellywilliam@ea-asia.com

Abstract

For companies, adhering to industry standards, regulations, and laws is essential. This includes the requirement to file annual financial statements using accounting standards issued by relevant government agencies. However, companies often document their transactions with various non-standard descriptions, leading to discrepancies between these transactions and the accounting standards. This project aims to address these discrepancies by proposing an automated tool that uses Artificial Intelligence (AI) and Machine Learning (ML) models to align company transactions with accounting standards. Eleven AI/ML models, including Random Forest (RF) and Decision Tree Classifier (DTC), were developed and trained using historical transactions. The evaluation results indicate that while both RF and DTC models have identical training accuracies, RF performs slightly better on the test set. Consequently, RF was chosen for the automated tool.

1 Introduction

For companies, compliance with industry standards, regulations, and laws is crucial (Lanturn Pte Ltd., 2023). For example, Singapore companies must file annual financial statements to the Accounting and Corporate Regulatory Authority (ACRA) (Accounting and Corporate Regulatory Authority, 2023). The main purpose of compliance is to identify and avoid red flags that may appear in the companies. All companies, including small companies, must comply with these regulatory requirements. Violations of regulatory compliance often result in legal punishment, including governmental fines (Accounting and Corporate Regulatory Authority, 2023). For small companies with limited resources, these compliances, including preparing and filing the annual financial statement, can be challenging (Wolters Kluwer, 2021).

Available accounting software (such as Xero (Xero Limited, 2023), Sage (Sage Group, 2023), QuickBooks (Intuit Inc., 2023), KashFlow (KashFlow Software Ltd., 2023), and MYOB (Asian Business Software Solutions Pte Ltd., 2023)) have been used to assist the companies to prepare their financial statement. This software helps in documenting, managing, and tracking a company's transactions. However, companies may have a lot of variations in their transactions. When the companies document their transactions using accounting software or manually, they may use various descriptions (non-standard descriptions). The government agencies have standard guidelines and classification for these accounts. Hence, they would need to manually map their transactions with the accounting standards. These processes require significant time and manpower and would require accounting knowledge. Depending on the size of the company and the complexity of the transactions, it may take around 2 days to 3 weeks to complete it.

This work aims to solve this problem by developing an automated tool to map the company's transactions with accounting standards. Powered by Artificial intelligence (AI) and Machine Learning (ML) models, the proposed tool would get the input in terms of accounting transactions and automatically map the transactions with the standard accounts. For this purpose, eleven AI/ML models are developed and trained using historical transactions. These models include Random Forest (RF) (Breiman, 2001), Balanced Random Forest (Kobylinski & Przepiorkowski, 2008), Decision Tree Classifier (DTC) (Suthaharan, 2016), Gradient Boosting (Sarkar & Natarajan, 2019), Multi-Layer Perceptron (MLP) (Manaswi, 2018), Ridge Classifier (Hastie, Tibshirani, & Friedman, 2009), Linear Support Vector Classifier (Linear SVC) (Bonaccorso, 2017), K-Nearest Neighbors (KNN) (Dangeti, 2017), Multiclass Logistic Regression (Li, Lin, & Zeng, 2023), Naive Bayes (NB) (Bonaccorso, 2017), and TensorFlow Keras Neural Network (NN) (Silaparasetty, 2020). The evaluation results show that RF and DTC models show identical training accuracies (0.74), but RF performs slightly better on the test set (0.60 vs. 0.54) at the cost of a longer runtime (136 seconds vs. 65 seconds). Based on these results, RF was selected to be used in the automated tool.

2 Data Pre-Processing

The proposed tool would get the input in terms of accounting transactions from the companies that have been prepared by using accounting software or manually prepared and map it to the accounting standards. The accounting standards are considered as a label for each transaction and AI/ML models would need to classify the transactions to the correct label. The accounting standards can have several levels as illustrated in Table 1. The maximum level of the accounting standards used in this work is 5 levels. However, not all transactions would have a complete 5 levels. For example, transaction #1 in Table 1 only has 4 levels, while transaction #5 has 5 levels.

To train AI/ML models, we use a historical dataset that contains past transactions from a local company in Singapore. The dataset is prepared manually, hence before applying the AI/ML models, the dataset undergoes data pre-processing to remove any data quality problems and transform the data input to fit the models. The steps in the data pre-processing are as follows:

1. Replace missing values. As mentioned earlier, each transaction may have different accounting levels. For a transaction that only has 1 level, 'Level 2' to 'Level 5' will be blank or missing. We replace this blank or missing in 'Level 1' to 'Level 5' with the value from the previous level. For example, transaction #1 in Table 1 only has 4 levels, 'Level 5' is blank. We replace 'Level 5' with 'Level 4' value. A sample of the dataset after the missing value replacement is in Table 2.

Table 1. Sample of the accounting standards for the company's transactions

No	Description	Level 1	Level 2	Level 3	Level 4	Level 5
1	TRADE DEBTORS	Assets	Current assets	Trade and other receivables	Trade Receivables	-
2	DEPOSIT & PREPAYMENT	Assets	Current assets	Trade and other receivables	Other receivables	Deposit
3	TERM LOAN	Assets	Current assets	Trade and other receivables	Other receivables	Related party
4	BANK (USD)	Assets	Current assets	Cash and bank balances	-	-
5	BANK (SGD)	Assets	Current assets	Cash and bank balances	-	-

Table 2. Sample of the dataset after the missing value replacement

No	Description	Level 1	Level 2	Level 3	Level 4	Level 5
1	TRADE DEBTORS	Assets	Current assets	Trade and other receivables	Trade Receivables	*Trade Receivables*
2	DEPOSIT & PREPAYMENT	Assets	Current assets	Trade and other receivables	Other receivables	Deposit
3	TERM LOAN	Assets	Current assets	Trade and other receivables	Other receivables	Related party
4	BANK (USD)	Assets	Current assets	Cash and bank balances	*Cash and bank balances*	*Cash and bank balances*
5	BANK (SGD)	Assets	Current assets	Cash and bank balances	*Cash and bank balances*	*Cash and bank balances*

2. Tag part-of-speech. To enhance our text data, we utilized the spaCy library (Explosion AI, 2024) to perform part-of-speech (POS) tagging. This method is used to analyze and annotate each word in the text with its corresponding part-of-speech category, such as noun, verb, and adjective. The processes of this tagging involve loading the English language model and processing each level of text data. A sample of the dataset after the tagging is in Table 3.

3. Lemmatizing the Text Data. We utilize a pre-existing lemma dictionary (Vidhya, 2024) to lemmatize our text data. This method is used to reduce inflected words to their base or dictionary form, which helps in normalizing text for analysis. This step involves replacing each word with its lemma based on its POS tag in the previous step. Common terms used in accounting are also added manually to the lemma dictionary. A sample of the lemma dictionary is in Table 4.

4. Feature engineering. After lemmatizing, we combined levels 1-5 into one column with a semicolon separator. Joining the multiple levels into a single column helps capture relationships between these levels which is important for the model to learn and reduces run time by reducing the number of labels to predict. An example of the cleaned dataset after the data pre-processing is shown in Table 5.

Table 3. Sample of the dataset after the part-of-speech tagging

No	Level 1	Level 2	Level 3	Level 4	Level 5
1	(assets, NOUN)	(current, ADJ) (assets, NOUN)	(trade, NOUN) (and, CCONJ) (other, ADJ) (receivables, NOUN)	(trade, NOUN) (receivables, NOUN)	(trade, NOUN) (receivables, NOUN)
2	(assets, NOUN)	(current, ADJ) (assets, NOUN)	(trade, NOUN) (and, CCONJ) (other, ADJ) (receivables, NOUN)	(other, ADJ) (receivables, NOUN)	(deposit, NOUN)
3	(assets, NOUN)	(current, ADJ) (assets, NOUN)	(trade, NOUN) (and, CCONJ) (other, ADJ) (receivables, NOUN)	(other, ADJ) (receivables, NOUN)	(related, VERB) (party, NOUN)
4	(assets, NOUN)	(current, ADJ) (assets, NOUN)	(trade, NOUN) (and, CCONJ) (other, ADJ) (receivables, NOUN)	(trade, NOUN) (receivables, NOUN)	(trade, NOUN) (receivables, NOUN)
5	(assets, NOUN)	(current, ADJ) (assets, NOUN)	(trade, NOUN) (and, CCONJ) (other, ADJ) (receivables, NOUN)	(other, ADJ) (receivables, NOUN)	(deposit, NOUN)

Table 4. Sample of the Lemma dictionary

No	Word	Dictionary
1	has	NOUN: has; VERB: have; AUX: have
2	insight	NOUN: insight
3	the	NOUN: the; PROPN: the; DET: the; PRON: the
4	mechanisms	NOUN: mechanism
5	learning	NOUN: learning; VERB: learn
6	and	NOUN: and; DET: a; CCONJ: and
7	processing	NOUN: processing; PROPN: Processing; VERB: process

Table 5. Sample of the dataset after cleaning

No	Description	Level 1	Level 2	Level 3	Level 4	Level 5	Combined Level
1	TRADE DEBTORS	Assets	Current assets	Trade and other receivables	Trade Receivables	*Trade Receivables*	asset;current asset;trade and other receivable;trade receivable;trade receivable
2	DEPOSIT & PREPAYMENT	Assets	Current assets	Trade and other receivables	Other receivables	Deposit	asset;current asset;trade and other receivable;other receivable;deposit
3	TERM LOAN	Assets	Current assets	Trade and other receivables	Other receivables	Related party	asset;current asset;trade and other receivable;other receivable;relate party

3 AI/ML Models

After the data preprocessing, we train AI/ML models using the cleaned dataset. We develop and train eleven AI/ML models. A short description of these eleven models is in Table 6.

Table 6. AI/ML models used

No	Model	Description	Example of use case(s)
1	Random Forest (RF)	RF is a supervised learning method based on a combination of decision tree predictors. Each tree depends on an independent random vector with the same distribution for all the forest trees (Breiman, 2001). It builds multiple decision trees using some component of randomness and determines the classification using most various trees.	Credit scoring to predict the likelihood of a borrower defaulting on a loan. Predicting customer churn in a telecom company by analyzing customer behavior and usage patterns.
2	Balanced Random Forest	Balanced Random Forest is an extension of the Random Forest algorithm designed to handle imbalanced datasets by under-sampling the majority class and over-sampling the minority class during the training process. Each tree will have bootstrapped sets of the same size, one for the minority class and the other for the majority class (Kobylinski & Przepiorkowski, 2008).	Medical diagnosis where the prevalence of a disease is low but critical to identify correctly. Detecting fraudulent transactions in financial datasets where fraudulent cases are much fewer than legitimate ones.
3	Decision Tree Classifier (DTC)	A Decision Tree Classifier is a non-parametric supervised learning method used for classification. It splits the data into subsets based on the value of input features, using a tree-like model of decisions (Suthaharan, 2016).	Customer churn prediction to identify customers likely to leave a service. Diagnosing diseases based on patient symptoms and medical history.
4	Gradient Boosting	Gradient Boosting is an ensemble technique that builds models sequentially, each correcting the errors of its predecessor (Sarkar & Natarajan, 2019).	Fraud detection in financial transactions. Predicting house prices by learning patterns from historical data.
5	Multi-Layer Perceptron (MLP	A Multi-Layer Perceptron (MLP) is a class of feedforward artificial neural networks. It consists of at least three layers of nodes: an input layer, a hidden layer, and an output layer. MLPs use backpropagation for training the network (Manaswi, 2018).	Image classification tasks Handwritten digit recognition.
6	Ridge Classifier	A Ridge Classifier is a type of linear classifier that applies L2 regularization (Hastie, Tibshirani, & Friedman, 2009) to prevent overfitting by penalizing large coefficients in the linear model.	Sentiment analysis of textual data. Text classification.

7	Linear Support Vector Classifier (Linear SVC)	Linear Support Vector Classifier (SVC) is a linear classification model that attempts to find the best hyperplane which separates data points of different classes with a maximum margin (Bonaccorso, 2017).	Spam email detection. Sentiment analysis to classify the sentiment of user reviews as positive or negative.
8	K-Nearest Neighbors (KNN)	K-Nearest Neighbors (KNN) is a non-parametric method used for classification and regression. It classifies a sample based on the majority class among its k-nearest neighbors in the feature space (Dangeti, 2017).	Recommender systems for suggesting products to users.
9	Multiclass Logistic Regression	Multiclass Logistic Regression extends binary logistic regression to handle multiclass classification problems (Li, Lin, & Zeng, 2023). While binary logistic regression is a conditional probability distribution.	Image classification with multiple categories. Classifying types of flowers based on their petal and sepal measurements.
10	Naive Bayes (NB)	Naive Bayes is a probabilistic classifier based on Bayes' theorem with the assumption of independence between features (Bonaccorso, Naive Bayes, 2017).	Document classification, such as classifying news articles.
11	TensorFlow Keras Neural Network (NN)	TensorFlow Keras is an open-source library for building neural networks. It provides an easy-to-use API for creating and training models, allowing for easy and fast prototyping of neural networks (Silaparasetty, 2020). The neural network finds underlying relationships in a dataset by applying a number of techniques.	Image recognition and classification in medical imaging. Predicting stock prices using a time series of historical data.

4 Results and Discussion

To evaluate the eleven AI/ML models, we use 5-fold cross-validation in the same environment and setup. The cutoff time to complete the training is 3 hours. The training accuracy and testing accuracy are calculated using the scikit-learn accuracy score (Scikit-learn , 2024). The result is shown in Table 7. The training accuracy, testing accuracy, and runtime shown in Table 7 are the average accuracy and runtime for each model.

From the result, several observations can be made. First, RF and DTC show identical training accuracies (0.74), but RF performs slightly better on the test set (0.60 vs. 0.54) at the cost of a longer runtime (136 seconds vs. 65 seconds). MLP achieves a high training accuracy (0.73) and testing accuracy (0.59), but it has a significantly longer runtime (2362 seconds). NB has the shortest runtime (2 seconds) but also the lowest performance (training accuracy 0.32, testing accuracy 0.31). Balanced RF significantly underperforms compared to the other models, with a training accuracy of 0.26 and a testing accuracy of 0.19. Gradient Boosting training is not completed after the cutoff time. Linear

SVC shows a balanced performance between accuracy (training accuracy 0.69, testing accuracy 0.56) and runtime (117 seconds). And TensorFlow Keras NN model performs comparably in testing accuracy (0.54) to DT (0.54) and RF (0.60) with a moderate runtime of 101 seconds. These observations indicate that RF is the best model for the mapping. Hence, RF was chosen for the automated tool.

Table 7. The performance of each model

Model	Training Accuracy	Testing Accuracy	Runtime (Sec)
RF	0.74	0.60	136
Balanced RF	0.26	0.19	47
DTC	0.74	0.54	65
Gradient Boosting	0.66**	0.45**	10,842**
MLP	0.73	0.59	2,362
Ridge Classifier	0.58	0.48	598
Linear SVC	0.69	0.56	117
KNN	0.59	0.47	89
Multiclass Logistic	0.49	0.44	906
NB	0.32	0.31	2
Tensorflow Keras NN	0.53	0.54	101

**: the training was not yet completed after the cutoff time

To further validate the mapping accuracy, we use the best model, RF, to map a new unlabeled dataset. The prediction result is then evaluated by two accounting experts. The experts confirm that the prediction is somewhat accurate with a minimum number of discrepancies between the predicted label and the accounting standard. Based on this result, we conclude that the model can map the transactions to the accounting standards.

5 Summary and Future Research

In this paper, we focus our study on developing an automated tool that uses Artificial Intelligence (AI) and Machine Learning (ML) models to align company transactions with accounting standards. For this purpose, eleven AI/ML models, including Random Forest (RF), are developed and evaluated using historical transactions. The evaluation using 5-fold cross-validation indicates that RF has the best training and testing accuracy with relatively short runtime. We then further validate the model by predicting the accounting standards for unmapped transactions and manually evaluating the predictions. The manual evaluation confirms that the prediction is somewhat accurate with a minimum number of discrepancies between the predicted label and the accounting standard.

As this automated tool development is still ongoing, we would like to continue improving the accuracy of the models by further fine-tuning the model using Reinforcement Learning from Human Feedback (RLHF) (Zhu, Jiao, & Jordan, 2023). We also would like to deploy and integrate the RF model with the web application to provide an easy interface for the users to upload the new data and map the transactions in the new data with the accounting standards.

Acknowledgement

This research is supported by the Temasek Polytechnic, under its Temasek Polytechnic Research Fund (TPRF) (15th Call).

References

Accounting and Corporate Regulatory Authority. (2023, July 27). *Singapore Financial Reporting Standard for Small Entities.* Retrieved from Accounting and Corporate Regulatory Authority: https://www.acra.gov.sg/accountancy/accounting-standards/pronouncements/singapore-financial-reporting-standard-for-small-entities

Asian Business Software Solutions Pte Ltd. (2023, August 28). *MYOB.* Retrieved from MYOB: https://sg.abssasia.com/free-trial?gclid=Cj0KCQjw6KunBhDxARIsAKFUGs_3kz5r6E3PbvwlQUL1hO6hdzxMkZXH-LvYTfRBN-XUoQY0x8RQr-EaArvjEALw_wcB

Bonaccorso, G. (2017). Naive Bayes. In *Machine Learning Algorithms.* UK: Packt Publishing.

Bonaccorso, G. (2017). *Support Vector Machines.* UK: Packt Publishing.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5-32.

Dangeti, P. (2017). K-Nearest Neighbors and Naive Bayes. In *Statistics for Machine Learning.* UK: Packt Publishing.

Explosion AI. (2024, April 24). *Linguistic Features.* Retrieved from spaCy: https://spacy.io/usage/linguistic-features

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* New York: Springer New York.

Intuit Inc. (2023, August 28). *Accounting Software | Quickbooks Singapore.* Retrieved from Quickbooks : https://quickbooks.intuit.com/sg/

KashFlow Software Ltd. (2023, August 28). *Kashflow Accounting Software.* Retrieved from Kashflow Accounting Software: https://www.kashflow.com/

Kobylinski, L., & Przepiorkowski, A. (2008). Definition Extraction with Balanced Random Forests. *Definition Extraction with Balanced Random Forests* (pp. 237–247). Berlin, Germany: ADVANCES IN NATURAL LANGUAGE PROCESSING.

Lanturn Pte Ltd. (2023, June 30). *Introduction to Singapore Regulatory Compliance Requirements*. (Lanturn Pte Ltd.) Retrieved September 10, 2024, from https://www.lanturn.com/sg/blog/introduction-to-regulatory-compliance-in-singapore

Li, H., Lin, L., & Zeng, H. (2023). Logistic Regression and Maximum Entropy Model. In *Machine Learning Methods* (pp. 103–125). Singapore: Springer.

Manaswi, N. K. (2018). Multilayer Perceptron. In *Deep Learning with Applications Using Python* (pp. 45–56). United States: Apress L. P.

Sage Group. (2023, August 28). *Business Tools and Advice from Real Expert.* Retrieved from Sage Group: https://www.sage.com/en-sg/

Sarkar, D., & Natarajan, V. (2019). Boosting Model Performance with Boosting. In *Ensemble Machine Learning Cookbook.* United Kingdom: Packt Publishing.

Scikit-learn . (2024, April 15). *Sklearn Metrics Accuracy Score.* Retrieved from Scikit-learn developers: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.accuracy_score.html

Silaparasetty, V. (2020). Neural Networks. In *Deep Learning Projects Using TensorFlow* (pp. 71–86). UK: Apress L.P.

Suthaharan, S. (2016). Machine Learning Models and Algorithms for Big Data Classification. In *Thinking with Examples for Effective Learning* (pp. 237–269). Netherlands: Springer Nature.

Vidhya, A. (2024, April 24). *How to Build a Lemmatizer.* Retrieved from Medium: https://medium.com/analytics-vidhya/how-to-build-a-lemmatizer-7aeff7a1208c

Wolters Kluwer. (2021, February 13). *What is an LLC annual report and how to file one for your business.* Retrieved from Wolters Kluwer: https://www.wolterskluwer.com/en/expert-insights/what-is-an-llc-annual-report-and-how-to-file-one-for-your-business

Xero Limited. (2023, August 28). *Accounting software to do your to-do.* Retrieved from Xero Limited: https://www.xero.com/

Zhu, B., Jiao, J., & Jordan, M. I. (2023). *Principled Reinforcement Learning with Human Feedback from Pairwise or K-wise Comparisons.* arXiv.