International Journal of Smart Sensor and Adhoc Network

Volume 3
Issue 4 *Role of Emerging & Intelligent Technologies for the Society.*
Article 2

January 2023

# Improvement of Key Financial Performance Indicators in the Insurance Industry Using Machine Learning – A Quantitative Analysis

Vineeth Jeppu
*NYIT*, vineethj007@hotmail.com

# Improvement of Financial KPIs in the Insurance Industry Using Machine Learning – A Quantitative Analysis

Vineeth Jeppu[1], Ayan Singh[2], Alex Gonzalez[3]

[1]New York Institute of Technology, NYC, New York
[2]San Jose State University, San Jose, California
[3]Rutgers University, New Brunswick, New Jersey
[1]*v*ineethj007@hotmail.com

***Abstract*—**AI and Machine learning are playing a vital role in the financial domain in predicting future growth and risk and identifyingkey performance areas.We look at how machine learning and artificial intelligence (AI) directly or indirectly alter financial management in the banking and insurance industries. First, a non-technical review of the prior machine learning and AI methodologies beneficial to KPI management is provided.This paper will analyze and improve key financial performance indicators in insurance using machine learning (ML) algorithms. Before applying an ML algorithm, we must determine the attributes directly impacting the business and target attributes. The details must be manually mapped from string values to fit the model and its required datatypes for applying these specific features to an ML model. We propose hashing to convert string values to numeric values for data analysis within our model. After the string values are hashed, we can introduce our model. In our case, we have chosen to use a decision tree model. Decision Trees are beneficial for this use case as this algorithm generates rulesets that govern the target value output. These rulesets can then be applied to the financial dataset and infer the "best fit" value that might be wrong/missing. Finally, because of the model, we can use this most accuratedata version to detect general ledger transactional data patterns.

***Keywords***—Machine Learning,Decision Tree, Hashing, Key Performance Indicator (KPIs), Insurance Metrics

## I. INTRODUCTION

The last two decades have seen a dramatic accelerating pace in the development and adoption of new technologies; however, rapid technological change can outpace the capacity of society to adapt [1]. On the other hand, we can see this overflow of new tech as an opportunity to improve upon your outdated legacy systems at your own pace. These obsolete systems include analysis via Excel, simple algorithms that need regular heavy supervision, SAS code, etc. This exploratory paper will discuss how to use a machine learning algorithm and how it can help a business improve its current functions and processes. We will also discuss the workflow for implementing a Machine Learning algorithm. The focus of our paper will be on using research and new concepts in ML to help assess and improve Key Performance Indicators (KPIs) in the Insurance Sector, as well as concluding on its benefits compared to the current workflow.

## II. BACKGROUND

KPIs, or Key Performance Indicators, are the real measures of organizational performance and well-being. Finding the correct ones and evaluating these KPIs are critical in inter and intra-company business performance statistics and sanctions. It is always very critical to understand which KPIs in marketing,finance, HR, and growth goals drive performance to chart a clear route to success [2]. As a result, AI is essential to recognizing and grouping such characteristics in order to advise the firm on the next step. In a complicated system, deciding which threads or processes have defects and how to fix them is extremely difficult. Analysts and managers do not have end-to-end access to all functions or sub-processes to check and advise. AI will automate many of these jobs or procedures by clearly communicating what needs to be rectified and how soon [3-5].Managers or analysts who see AI as a collaborator will see there is no need to "race against a computer." While human judgment is unlikely to be automated, intelligent robots may significantly contribute to this job by aiding with decision support and data-driven simulations, among other things that will impact the KPIs.

KPIs are the critical metrics that monitor to have the most influence on strategic business results. KPIs help the team focus on what's essential and support the long-term organizational plan. KPIs are developed utilizing the business SMART framework (Specific, Measure, Attainable, Relevant, Timeframe) [6-8]. KPIs must be defined based on the following criteria [9] as described below.

- Is your goal specific?
- Can you quantify your progress toward your goal?

- Is the objective practically attainable?
- How vital is the aim of your company?
- What is the timetable for accomplishing this objective?

On the other hand, KPIs may be used to measure the success of AI initiatives. Organizations should correctly define KPIs to increase project efficiency while enabling them to help society [7, 10, 11]. This job is for both achievements: measuring organizational and AI KPIs in a particular setup.

## III. DATASET DESCRIPTION

Data has always been at the heart of the business. Data is an essential component of every AI model and, in essence, the fundamental reason for machine learning's present popularity. Scalable ML algorithms have become practical solutions that may bring value to a business rather than being a consequence of its core activity due to the availability of data [12].

The dataset included two months of loss reconciliation data. It has a few parameters like paidLoss, caseReserve, and so on. It included 1 million rows and 500 fields. The data was unprocessed and needed to be treated before it could be used to train the model. The figure below (see Fig 1) shows the data splitting in a standard ML modeling setup [13-15].

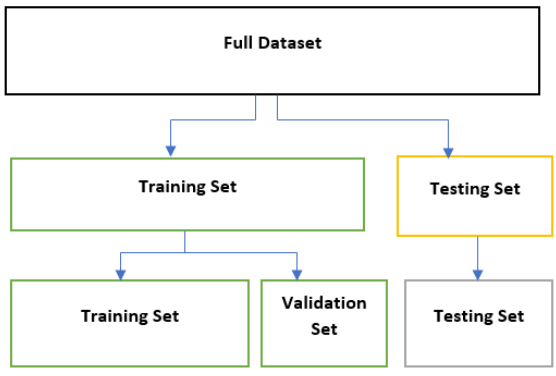


**Fig 1**. Splitting of data into training, testing, and validation datasets.

## IV. IMPLEMENTATION AND METHODOLOGY

Before Machine learning was introduced and its applications were developed, trends in data and other valuable information were extracted through statistics or with the help of spreadsheets [15, 16]. These statistics were then used to boost a company's business and provide insights into its financials.

In the present, however, Machine learning Algorithms can be used to interpret the trends more accurately in the data and then turn those trends or patterns into business uses. Supervised learning, the most used machine learning, helps predict and classify data [3,14,16]. The insurance and financial industries use Supervised algorithms such as classification, where the data is divided into individual attributes and features which correspond to one of the target or output variables [17].

Our chosen model, decision trees, falls under the classification model category of supervised learning programs. The output of the decision tree will be a ruleset that can be used to "classify" the input variables/attributes under a particular target.[6] For example, when a financial institutionhas to categorize a business resource, they will use classification where in which it is found to belong to one of the business units via the predetermined ruleset [18].

For our cases, we intend to take the list of target attributes, three to four different points, and create their rulesets in which new data can be applied. With these rulesets, we intend to make a function in which the model can use them for a new data set. This will help correct any entries and fill in any blanks in the dataset by predicting the "best fit" value based on the historical rulesets [6, 19, 20]. This will help keep data storage accurate and help the business analyst make informed decisions about the company. Speaking generally, our model can predict and fill in the values for Business units, product codes, and various other attributes. With your new data version, you can create visualizations and make forecasts/predictions of the actual trajectory of the business units with a smaller margin of error [17, 21]. As a result, you can be more confident in key performance indicators, such as profit margin, revenue growth, and client retention rate.

The basic workflow of an ML model can be understood with the help of a workflow diagram below [22]. The steps that need to be followed are as follows:

Collect - When looking for a dataset, the first step is to decide on the sources from which you will acquire the data.

Preprocess - Every skilled data scientist follows a concept todetermine if a dataset is new or previously utilized.

Annotate - After ensuring that your data is clean and useful, you must guarantee that it is intelligible by a machine.

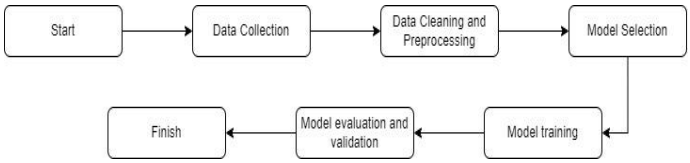


**Fig 2**. Basic ML workflow in KPI Metric analysis

Further analysis must be made to draw such conclusions; however, our focus was on creating and improving upon a process in which critical business attributes in data sets can be logically predicted.

## V. MODEL WORKFLOW

### *A. Data Collection*

As a part of every machine learning project, we need data to develop the model. The more data we have, the better we can train the model.

### *B. Data Preparation*

(a) Data Cleaning: Data cleaning is known as 'data pre-processing.' Data Cleaning involves removing outliers and corrupt and incomplete records from the dataset. Some steps followed, replacing the missing value with the mean, beginning, etc.

(b) Exploratory Data Analysis (EDA): EDA is a vital part of any ML model to have the right inputs feed into the model. In EDA, we try to understand the data before we try to fit any machine learning model into the data. This is done through visualization of graphs, determination, or running experiments like analyzing correlations between potential features and target variables. Visualizations like histograms also assist with selecting feature attributes that have unique values under a certain threshold. After studying data, we can adjust the amount of information gained by each feature chosen attribute by adjusting hyperparameters. This is particularly useful in classification problems this paper aims to solve.

(c) Data Transformation: The attributes must be mapped from string values to fit the model and its required datatypes. We propose hashing to convert string values to numeric values for data analysis within our model.

### *C. Model Selection*

Once data preparation is finished, we need to decide the model we need to choose for the training. Understanding the type of problem at hand will help in model selection. For example, models such as Naïve Bayes, Decision Tree, Random Forest, and K-Nearest neighbor are popular in solving classification problems [18, 23].

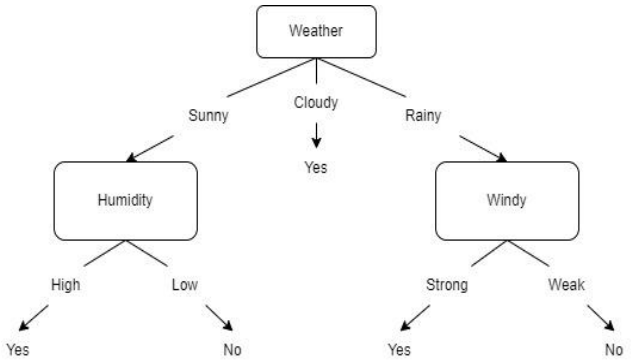


**Fig 3**. Decision tree

## VI. MODEL TRAINING AND VALIDATION

After the model selection, the research team used the training data or trained the model. The training data is created from a train-test split or can be a separate dataset on which the model is introduced.

Once the model is trained, the evaluation of the model will help us understandits accuracy of the model. The review of the model is done using the test dataset. We used formulas to evaluate the model's accuracy. We made the following observations: we obtained an accuracy score of 99.79 for the first model, 93.79 for the second model, and 97.94 for the third. Three different models were built according to three other KPI variables (see Table 1).

**TABLE 1**. MODEL OUTPUTS BY ITERATION

| Training Model (Decision Tree) | Accuracy Score |
|---|---|

| | |
|---|---|
| Iteration 1 | 99.79 |
| Iteration 2 | 93.79 |
| Iteration 3 | 97.94 |

## VII. RESEARCH TAKEAWAYS

In this process, we learned how to effectively work as a team on one project across departments to collect the right inputs. We learned more about machine learning algorithms, including their requirements and how to implement them effectively. We learned how to hash string values so they can be used in the model and then provide a lookup key to translate the hashed value for the predictions outputted by the model. The team's significant takeaways are the importance of maturing a model and focusing on a project's technical and business requirements. All units are encouraged to take on research projects and proactively find new and exciting things to improve productivity and ease of use.

## VIII. RECOMMENDATIONS BASIS OF DATA

The details must be manually mapped from string values to fit the model and its required datatypes for applying these specific attributes to a machine-learning model.We propose hashing to convert string values to numeric values for data analysis within our model. Hashing enables the mapping of different forms of data to fixed-sized integer values. Hashing is used to store data using lookup values in dictionaries while minimizing "collisions," i.e., two data sets have the same hash value. Python uses the in-built Sip-Hash function for hashing [8, 18] as these data have PII data and must be handled cautiously. Our approach is to hash string values to integers of order eight and then use the hashed values as inputs for our Decision Tree model.

## IX. CONCLUSIONS

These decision trees can perform missing value imputations on our specification's target value(s). This is important because accurately predicting missing values on the general ledger can help improve data insights and create the most accurate financial data. Decision Trees are beneficial for this use case as this algorithm generates rulesets that govern the target value output. These rulesets can then be applied to the financial dataset and infer the "best fit" value that might be wrong/missing. Finally, this most accurate data version can detect general ledger transactional data patterns. This will help correctly assign dollar value amounts where our model-recognized cash values have been incorrectly set.This way, you can assess the business and make better forecasts based on the most accurate data.


## ACKNOWLEDGMENT

We want to thank our manager Dimple Prajapatiand the rest of the QBE team for giving us opportunities as interns, providing us the opportunity to perform this research, and supporting our decisions for our work.


## REFERENCE


[1]T Mitchell,, B Buchanan,, G DeJong,, T Dietterich,, P Rosenbloom, and, and A Waibel Annual Review of Computer Science 1990 4:1, 417-433 https://www.annualreviews.org/doi/abs/10.1146/annurev.cs.04.060190.002221?journalCode=arcompsci

[2] Kolbjørnsrud, V., Amico, R., & Thomas, R. J. (2016). How artificial intelligence will redefine management. Harvard business review, 2(1), 3-10.

[3] Rentala, V. K., Kanzler, D., & Fuchs, P. (2022). POD Evaluation: The Key Performance Indicator for NDE 4.0. Journal of Nondestructive Evaluation, 41(1), 1-12.

[4] Stevens, N. A., Lydon, M., Marshall, A. H., & Taylor, S. (2020). Identification of bridge key performance indicators using survival analysis for future network-wide structural health monitoring. Sensors, 20(23), 6894.

[5] Gao, P., Li, J., & Liu, S. (2021). An introduction to key technology in artificial intelligence and big data driven e-learning and e-education. Mobile Networks and Applications, 26(5), 2123-2126.

[6] Van de Vrande, V., De Jong, J. P., Vanhaverbeke, W., & De Rochemont, M. (2009).

Open innovation in SMEs: Trends, motives and management challenges. Technovation, 29(6-7), 423-437.

[7] Britannica, T. Editors of Encyclopaedia (2021, May 6). *algorithm*. *Encyclopedia Britannica*. https://www.britannica.com/science/algorithm

[8] IBM. (2017). Decision tree models. IBM. Retrieved August 23, 2022, from https://www.ibm.com/docs/en/spss-modeler/18.1.1?topic=trees-decision-tree-models

[9] United Nations Conference on Trade and Development (Unctad), THE IMPACT OF RAPID TECHNOLOGICAL CHANGE ON SUSTAINABLE DEVELOPMENT. (n.d.). https://unctad.org/system/files/official-document/dtlstict2019d10_en.pd

[10] Rawat, S., Rawat, A., Kumar, D., &Sabitha, A. S. (2021, April 2). Application of machine learning and data visualization techniques for decision support in the insurance sector. International Journal of Information Management Data Insights. Retrieved August 9, 2022, from https://www.sciencedirect.com/science/article/pii/S2667096821000057

[11] Miller, B., &Ranum, D. (2018). 5.5. hashing. 5.5. Hashing –Problem-Solving with Algorithms and Data Structures. Retrieved August 23, 2022, from http://www.openbookproject.net/books/pythonds/SortSearch/Hashing.html

[12] Dash, B. (2021). A hybrid solution for extracting information from unstructured data using optical character recognition (OCR) with natural language processing (NLP).

[13] Dash, B., Sharma, P., & Ali, A. (2022). Federated learning for privacy-preserving: A review of PII data analysis in Fintech. International Journal of Software Engineering & Applications, 13(4), 1-13.

[14] Kiron, D. (2022). AI Can Change How You Measure-and How You Manage. *MIT Sloan Management Review*, *63*(3), 24-28.

[15] Ansari, M. F., Dash, B., & Sharma, P. (2022). The Impact and Limitations of Artificial Intelligence in Cybersecurity: A Literature Review.

[16] Novillo, C. S., Crespo, A. B., Brualla, A., Bergelson, N., Hickman, C., & Márquez, C. (2022). Automatic assessment of Time-Lapse videos using CHLOE-EQ can automate KPI assessment to validate the operational performance of an IVF Clinic. *Reproductive BioMedicine Online*, *45*, e3-e4.

[17] Dash, B., Sharma, P., Ansari, M. F., & Swayamsiddha, S. (2022). A review of ONDC's digital warfare in India taking on the e-commerce giants. *International Journal of Engineering & Technology*, *11*(2), 96-99.

[18] Deußer, T., Ali, S. M., Hillebrand, L., Nurchalifah, D., Jacob, B., Bauckhage, C., &Sifa, R. (2022). KPI-EDGAR: A Novel Dataset and Accompanying Metric for Relation Extraction from Financial Documents. *arXiv preprint arXiv:2210.09163*.

[19] Kiron, D. (2022). AI Can Change How You Measure-and How You Manage. MIT Sloan Management Review, 63(3), 24-28.

[20] Olan, F., Arakpogun, E. O., Suklan, J., Nakpodia, F., Damij, N., &Jayawickrama, U. (2022). Artificial intelligence and knowledge sharing: Contributing factors to organizational performance. Journal of Business Research, 145, 605-615.

[21] Maschke, R. W., Pretzner, B., John, G. T., Herwig, C., &Eibl, D. (2022). Improved Time Resolved KPI and Strain Characterization of Multiple Hosts in Shake Flasks Using Advanced Online Analytics and Data Science. Bioengineering, 9(8), 339.

[22]Dash, B., & Ansari, M. F. (2022). Self-service analytics for data-driven decision making during COVID-19 pandemic: An organization's best defense. Academia Letters, 2.

[23] Saleheen, F., & Habib, M. M. (2022). Constructing Performance Measurement Index for

the Manufacturing Industry: An Empirical Study. Open Journal of Social Sciences, 10(3), 137-151.

**Biography**

**Vineeth Jeppu is** a second-year graduate student pursuing a Master of Science in Computer Science from the New York Institute of Technology. He received his bachelor's degree in Computer Science from NMAM Institute of Technology, Karnataka, India. He is passionate about the applications of machine learning to help humankind's needs. He is also interested in software engineering and application development.

**Ayan A. Singh** is a second-year graduate student enrolled in the Master of Science in Computer Science program at San Jose State University. He received his bachelor's degree (also in Computer Science) from Purdue University. He is passionate about applying deep learning algorithms to solve accessibility problems related to video games. He is also interested in software engineering and data science and grappling with issues that don't currently have adequate solutions.

**Alexandro (Alex) Gonzalez** is a fourth-year undergraduate student pursuing a Bachelor of Arts in Economics from Rutgers University - the State University of New Jersey (minor in Business Administration.) He is passionate about using and researching new technologies to help improve or replace current legacy systems. I am interested in project management and finance.