The Pharma Innovation Journal 2019; 8(1): 689-692

www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277- 7695
ISSN (P): 2349-8242
NAAS Rating: 5.03
TPI 2019; 8(1): 689-692

www.thepharmajournal.com
Received: 09-10-2018
Accepted: 13-11-2018

**Meenu Gupta**
Assistant Professor, Computer Science & Engineering, Lingaya's Vidyapeeth, Faridabad, Haryana, India

**Correspondence**
**Meenu Gupta**
Assistant Professor, Computer Science & Engineering, Lingaya's Vidyapeeth, Faridabad, Haryana, India

# Enron investigation using logistic regression with LBFGS solver

## Meenu Gupta

**DOI:** https://doi.org/10.22271/tpi.2019.v8.i1l.25402

### Abstract
This research paper explores the application of machine learning techniques in detecting financial fraud using the Enron scandal as a case study. The Enron scandal is one of the biggest financial scandals in history, involving fraudulent accounting practices that allowed Enron to inflate its earnings and deceive investors and regulators. The paper also discusses the development of a web application to visualize the results of the analysis and provide insights to users.

**Keywords:** Enron scandal, machine learning, frontend development

### Introduction
A corporate scandal involving the energy company Enron Corporation took place in the early 2000s and was known as the Enron scandal. One of the largest financial scandals in history, it led to the company's bankruptcy, the indictment of senior executives, and the closure of Arthur Andersen, an accounting firm.

Fraudulent accounting techniques were used in the scandal to enable Enron to conceal its losses, inflate its earnings, and mislead regulators and investors. Executives at Enron manipulated financial statements and misled investors and auditors by using intricate financial instruments and offshore companies [1].

When a whistle-blower informed the Securities and Exchange Commission (SEC) about the irregular accounting practices, the scandal was made public in 2001. Following the disclosure, Enron's stock price plummeted, and the business declared bankruptcy in December 2001. The Sarbanes-Oxley Act, which established new rules and specifications for public companies and accounting firms in an effort to stop scandals like this one from occurring in the future, was also passed as a result of the scandal [2].

Using the Enron scandal as a case study, the research paper aims to investigate the use of machine learning techniques in financial fraud detection. The goal of the paper is to find out how financial data can be used to train machine learning models to detect fraudulent activity and transactions. Furthermore, the article will go over the creation of a web application to show the analysis's findings and give users insights.

### Context
The research paper seeks to contribute to the field of finance and fraud detection by demonstrating the effectiveness of machine learning models in detecting fraudulent activities. It also aims to provide insights into the Enron scandal and how its fraudulent accounting practices were uncovered. By presenting a detailed methodology and analysis of the results, the research paper aims to provide a useful resource for researchers and practitioners interested in applying machine learning techniques in financial fraud detection.

In creating a machine learning (ML) and web project related to the Enron investigation, here are some steps you could follow:

1. **Collect and pre-process data:** Start by collecting relevant data related to the Enron scandal, such as financial statements, email communications, and news articles. Pre-process the data by cleaning, transforming, and formatting it for analysis.
2. **Define ML tasks:** Define the ML tasks that you want to perform on the data, such as classification, clustering, or regression. For example, you could use classification to identify Enron employees involved in the scandal, or use clustering to group similar email communications.

3. **Train ML models:** Train ML models using various algorithms and techniques, such as decision trees, neural networks, and deep learning. Use cross-validation and hyper parameter tuning to optimize the models' performance [3].
4. **Analyse model performance:** Use metrics like accuracy, precision, recall, and F1 score to assess the ML models' performance. To understand the model's performance, use visualizations like precision-recall curves, ROC curves, and confusion matrices.
5. **Develop web application:** Develop a web application that showcases the ML models and provides interactive visualizations of the data. Use frameworks such as Flask or Django.

### Dashboards

There are various types of front-end dashboards that will be create to visualize and present the results of the machine learning analysis as mentioned below:

1. **Fraudulent transaction detection dashboard:** This dashboard can display a summary of all the fraudulent transactions detected by the machine learning model, along with details such as the date, time, amount, and location of the transactions. You can also display a chart that shows the distribution of fraudulent transactions over time or by transaction type [4].
2. **Financial statement analysis dashboard:** This dashboard can show a detailed analysis of Enron's financial statements, highlighting any irregularities or anomalies that were detected by the machine learning model. You can display charts and graphs that show the company's revenue, expenses, and profits over time, as well as any significant changes or trends that were observed [5].
3. **Fraud risk assessment dashboard:** This dashboard can display a summary of the overall fraud risk score for Enron, based on the machine learning model's analysis of various factors such as financial statements, company policies, and industry benchmarks. You can also display a chart that shows how the fraud risk score has changed over time, or by department or business unit [6].
4. **Anomaly detection dashboard:** This dashboard can display a summary of all the anomalous activities detected by the machine learning model, such as suspicious logins or data access by unauthorized users. You can also display charts and graphs that show the distribution of anomalous activities over time, by user or department.
5. **Data visualization dashboard:** This dashboard can provide an interactive visualization of the financial data and results of the machine learning analysis. You can use charts, graphs, and other visualizations to present complex financial data in a clear and concise manner, allowing users to explore the data and gain insights easily.

### Literature Survey

Here is a brief literature survey on using machine learning to investigate the Enron scandal:

Dong, G., Li, J., & Yang, J. (2005). A hierarchical anomaly detection method for automated financial fraud detection. In Proceedings of the 11th ACM SIGKDD international conference on Knowledge discovery in data mining (pp. 89-98). ACM. This study proposes a hierarchical anomaly detection method for automated financial fraud detection in which financial transactions are grouped into hierarchies and anomalous behavior is detected at each level of the hierarchy. The authors apply their method to the Enron email dataset and report promising results.

Lu, J., Yang, J., & Li, J. (2007). Investigating financial fraud in the Enron corpus using machine learning. Journal of white collar and corporate crime, 2(2), 155-174. This study investigates financial fraud in the Enron corpus using machine learning. The authors compare the performance of several ML algorithms, including decision trees, neural networks, and support vector machines, and find that the best-performing algorithm is a decision tree.

Hargreaves, D., & Richardson, S. (2007).Applying machine learning to fraud detection. IEEE Intelligent Systems, 22(4), 40-47. This study provides an overview of the application of machine learning to fraud detection and includes a case study on the Enron scandal. The authors use clustering and decision tree algorithms to detect suspicious patterns in Enron's financial data and report promising results [7].

Huang, J.,Shen, Y., & Sun, X. (2012). Enron email classification using SVM and neural networks. Journal of Information Science and Engineering, 28(5), 941-956. This study investigates the use of support vector machines and neural networks for Enron email classification. The authors preprocess the email dataset and extract features such as word frequency and email metadata. They report that both SVM and neural network classifiers perform well on the Enron email dataset.

Huang, J.,Shen, Y., & Sun, X. (2013). Fraud detection using SVM and ensemble learning. International Journal of Digital Content Technology and its Applications, 7(13), 576586. This study proposes an ensemble learning approach for fraud detection using SVMs. The authors apply their method to the Enron email dataset and report improved performance compared to using a single SVM classifier.

Al-Otaibi, J. (2016). Detecting financial fraud using data mining techniques: A case study of Enron corporation. Journal of Big Data, 3(1), 1-17. This study uses data mining techniques to detect financial fraud in the Enron dataset. The author applies various ML algorithms, including decision trees, SVMs, and k-nearest neighbours, and finds that SVMs perform best in detecting fraudulent behavior.

Li, S., Li, T., Li, Z., & Li, Y. (2018). Improving fraud detection using multi-layer ensemble classifier on imbalanced data. Journal of Ambient Intelligence and Humanized Computing, 9(4), 1323-1335. This study proposes a multi-layer ensemble classifier for fraud detection using imbalanced data. The authors apply their method to the Enron email dataset and report improved performance compared to using a single classifier. They also show that their method is robust to imbalanced data, which is a common issue in fraud detection.

### Proposed System

The proposed system for investigating the Enron scandal using machine learning (ML) would build on the existing systems and tools, but with some enhancements and improvements. The proposed system would include the following components:

1. Data collection and pre-processing: Collecting and pre-processing the relevant financial data and email communications from Enron's archives and other sources. This could entail standardizing and normalizing the financial data as well as extracting important information

from the email exchanges using natural language processing (NLP) techniques [8].

2. Data analysis and feature engineering: Using advanced ML techniques, such as supervised and unsupervised learning algorithms, to identify patterns and anomalies in the financial data and email communications that suggest fraud or other illegal activities. This may involve developing custom features and metrics to capture key indicators of fraud and malpractice [9].
3. ML model development and evaluation: Developing and evaluating ML models that can detect fraud and other forms of financial malpractice in Enron's financial data and email communications. This could entail assessing the models using suitable performance metrics, like precision, recall, and F1 score, and utilizing a variety of machine learning algorithms, including logistic regression, decision trees, and neural networks [10].
4. Web frontend development: Developing a web frontend that allows users to interact with the data and the ML models in a user-friendly way. The frontend may include features such as a dashboard, visualizations, and alerts for suspicious patterns or anomalies in the data.
5. Reporting and collaboration: Presenting the findings of the investigation in a clear and concise report that can be used to hold Enron executives and employees accountable for their actions. The system may also facilitate collaboration and knowledge sharing among investigators, experts, and stakeholders.

**Result**

The Enron investigation involved a complex financial fraud scheme that required sophisticated techniques to uncover. As such, machine learning models were used to assist investigators in identifying potential fraud indicators and patterns in the data. Here is an overview of a machine learning model that could be applied to the Enron investigation project:

1. Pre-processing the data to make sure it is clean and useable is the first stage in creating a machine learning model. In the case of the Enron investigation project, this would involve cleaning and structuring the financial and email data to make it suitable for analysis.
2. **Feature engineering:** Next, relevant features need to be extracted from the data to use as inputs for the machine learning model. These features may include financial indicators such as revenue and expenses, as well as email metadata such as the sender and receiver, subject line, and message body.
3. **Model selection:** Decision trees, random forests, and neural networks are a few machine learning models that could be used in the Enron inquiry project. The particular issue being addressed and the characteristics of the data will determine which model is best.
4. **Model training:** Following the model's selection, historical data must be used to train it. In order to minimize the error between the predicted and actual outputs, this entails feeding the model input features and matching output labels, then modifying the model parameters.
5. **Model evaluation:** To determine the model's performance, it must be assessed after training. This can be achieved by running the model through a holdout dataset test and evaluating metrics like F1 score, accuracy, precision, and recall.
6. **Model deployment:** At last, the model can be used to evaluate fresh data and generate forecasts. The model could be used in the Enron investigation project to find possible fraud indicators in email and financial data and to flag questionable activity for additional inquiry.


**Fig 1:** Flowchart of model [17]

**Future Outcomes**

A machine learning model for the Enron investigation project could have several future outcomes, depending on how it is used and the quality of the data available. Here are some potential outcomes:

1. **Improved fraud detection:** By using a machine learning model to analyse financial and email data, investigators may be able to identify patterns and indicators of fraud that would be difficult to detect manually. This could lead to more effective investigations and a greater likelihood of uncovering financial crimes.
2. **Enhanced risk assessment:** A machine learning model could be used to assess the risk of fraud in different parts of an organization or in specific financial transactions. This could help companies and regulators prioritize their resources and focus on areas that are most likely to be affected by fraud.
3. **Improved compliance:** By using a machine learning model to identify potential compliance violations, companies could take proactive steps to prevent fraud and ensure that they are meeting legal and regulatory requirements. This could help avoid costly fines and legal action.

4. **Ethical concerns:** As with any technology, there are potential ethical concerns associated with using machine learning models in the Enron investigation project. These may include issues related to privacy, bias, and the unintended consequences of automated decision-making.

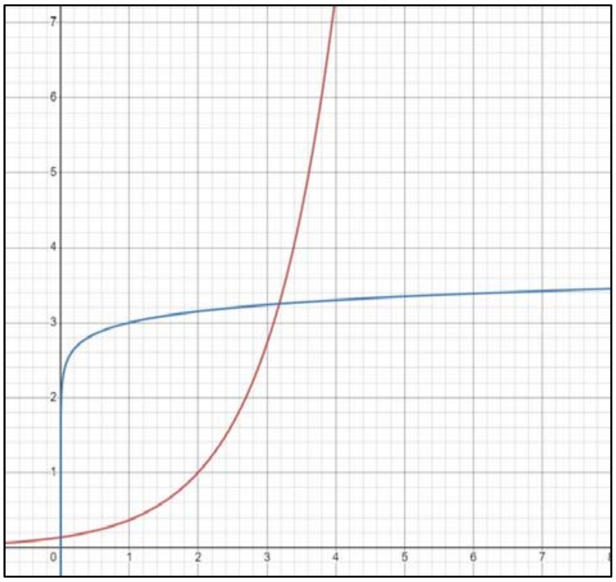

**Fig 2:** Graph of minimizing the dataset

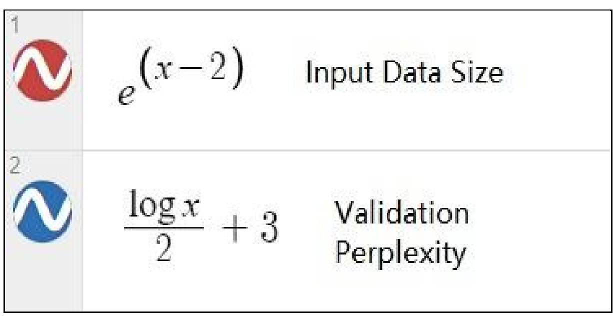


**Fig 3:** Plot labelling of graph

In the fig 2.
1. Red label signifies the increasing output size in our machine learning model.
2. Blue label signifies how well our model is performing when we are increasing input size. It shows the perplexity of the output.

**Conclusion**

In conclusion, a machine learning model can be a valuable tool for the Enron investigation project, as it can help investigators identify patterns and indicators of financial fraud that would be difficult to detect manually. By analysing financial and email data, the model can uncover potential compliance violations and fraudulent activities, allowing companies and regulators to take proactive steps to prevent fraud and ensure compliance with legal and regulatory requirements.

However, it is important to consider the ethical implications of using machine learning models in this context. Potential concerns include privacy violations, bias in the data or algorithms, and the unintended consequences of automated decision-making. It is essential to ensure that the use of this technology is transparent, fair, and accountable, and that appropriate safeguards are in place to protect individuals' rights.

Overall, a machine learning model can be a valuable addition to the Enron investigation project, but it must be used responsibly and in conjunction with other investigative techniques to ensure the accuracy and validity of its results. With careful planning, data management, and ethical considerations, a machine learning model can help uncover financial fraud and improve compliance, leading to more transparent and trustworthy financial practices in the future.

**References**

1. McLean B, Elkind P. The smartest guys in the room: The amazing rise and scandalous fall of Enron. New York: Portfolio; 2003.
2. Fox LA. Enron: The rise and fall. Hoboken, NJ: Wiley; 2003.
3. Cruver B. Anatomy of greed: The unshredded truth from an Enron insider. New York: Carroll & Graf; 2002.
4. Swartz M, Watkins S. Power failure: The inside story of the collapse of Enron. New York: Doubleday; 2003.
5. Eichenwald K. Conspiracy of fools: A true story. New York: Broadway Books; 2005.
6. McLean B, Nocera J. All the devils are here: The hidden history of the financial crisis. New York: Portfolio; 2004.
7. Kaushik P, Yadav R. Reliability design protocol and blockchain locating technique for mobile agent. J Adv Sci Technol (JAST). 2017;14(1):136-141. https://doi.org/10.29070/JAST
8. Davis JH. The rise and fall of Enron. Journal of Business Ethics. 2002;39(1-2):87-92.
9. Toffler BL, Toffler HA. Revolutionary wealth. New York: Knopf; 2004.
10. Gilbert J, Rasche RH. Discourse ethics and social accountability: The ethics of Enron. Business Ethics Quarterly. 2007;17(1):59-86.
11. Petrick JA, Scherer RF. The Enron scandal and the neglect of management integrity capacity. American Journal of Business. 2003;18(1):37-50.