

# Detecting Fraud and Misstatements in Annual Financial Statement Reports

Ryan Huang

Department of Computer Science  
Department of Economics  
University of North Carolina at Chapel-Hill  
Chapel Hill, USA  
rhuang1@unc.edu

Chunjiang Zhu

Department of Computer Science  
UNC Greensboro  
Greensboro, USA  
c\_zhu3@uncg.edu

Qianqian Tong

Department of Computer Science  
UNC Greensboro  
Greensboro, USA  
q\_tong@uncg.edu

**Abstract**—We develop a new fraud detection model that builds upon existing machine learning methods and incorporates recent technological advances in the Large Language Model field such as GPT-4 and GPT-4o. We demonstrate that combining deep learning knowledge with domain knowledge is beneficial towards fraud detection. We differ from previous research by using textual and linguistic data from annual 10-k financial statements in addition to the raw accounting numbers.

**Index Terms**—Fraud Detection, Large Language Models, Neural Networks, Machine Learning

## I. INTRODUCTION

This document is a model and instructions for L<sup>A</sup>T<sub>E</sub>X. Please observe the conference page limits.

## II. BACKGROUND

Filler text for now.

## III. MOTIVATION

The IEEEtran class file is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

## IV. METHODOLOGY

We use

### A. Machine Learning Methods

- 1) *Logistic Regression*:
- 2) *Random Forest*:
- 3) *SVM*:
- 4) *XGBoost*:
- 5) *RUSBoost*:

### B. Deep Learning Methods

- 1) *Hierarchical Attention Network (HAN)*:
- 2) *Long Short-Term Memory (LSTM)*:
- 3) *Multi-layer Perceptron (MLP)*:

### C. Large-Language Models (LLMs)

- 1) *GPT-4*:
- 2) *FinBERT*:
- 3) *LLaMA3*:

## V. DATA

Before you begin to format your paper, first write and save the content as a separate text file. Complete all content and organizational editing before formatting. Please note sections V-A–V-E below for more information on proofreading, spelling and grammar.

Keep your text and graphic files separate until after the text has been formatted and styled. Do not number text heads—L<sup>A</sup>T<sub>E</sub>X will do that for you.

### A. Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, ac, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

### B. Units

- Use either SI (MKS) or CGS as primary units. (SI units are encouraged.) English units may be used as secondary units (in parentheses). An exception would be the use of English units as identifiers in trade, such as “3.5-inch disk drive”.
- Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.
- Do not mix complete spellings and abbreviations of units: “Wb/m<sup>2</sup>” or “webers per square meter”, not “webers/m<sup>2</sup>”. Spell out units when they appear in text: “. . . a few henries”, not “. . . a few H”.
- Use a zero before decimal points: “0.25”, not “.25”. Use “cm<sup>3</sup>”, not “cc”).

TABLE I  
FINANCIAL VARIABLES

Abbreviation	Financial Variable
ACT	Total Current Assets
AT	Total Assets
AP	Accounts Payable
CEQ	Common Equity
CHE	Cash
COGS	Cost of Goods Sold
CSHO	Common Shares Outstanding
DLC	Current Debt
DLTT	Long Term Debt
DP	Depreciation
IB	Income Before Extraordinary Items
INVT	Inventories
IVAO	Investments Advances Other
IVST	Short Term Investments
LCT	Current Liabilities
LT	Total Liabilities
NI	Net Income
PPEGT	Gross Property Plant Equipment
PSTK	Preferred Stock
RE	Retained Earnings
RECT	Receivables
SALE	Sales
SEQ	Stockholders Equity
TXT	Income Tax
XINT	Interest Expense

### C. Equations

Number equations consecutively. To make your equations more compact, you may use the solidus ( / ), the exp function, or appropriate exponents. Italicize Roman symbols for quantities and variables, but not Greek symbols. Use a long dash rather than a hyphen for a minus sign. Punctuate equations with commas or periods when they are part of a sentence, as in:

$$a + b = \gamma \quad (1)$$

Be sure that the symbols in your equation have been defined before or immediately following the equation. Use “(1)”, not “Eq. (1)” or “equation (1)”, except at the beginning of a sentence: “Equation (1) is . . .”

### D. Challenges

Please use “soft” (e.g., `\eqref{Eq}`) cross references instead of “hard” references (e.g., (1)). That will make it possible to combine sections, add equations, or change the order of figures or citations without having to go through the file line by line.

Please don’t use the `{eqnarray}` equation environment. Use `{align}` or `{IEEEeqnarray}` instead. The `{eqnarray}` environment leaves unsightly spaces around relation symbols.

Please note that the `{subequations}` environment in  $\LaTeX$  will increment the main equation counter even when there are no equation numbers displayed. If you forget that, you might write an article in which the equation numbers skip from (17) to (20), causing the copy editors to wonder if you’ve discovered a new method of counting.

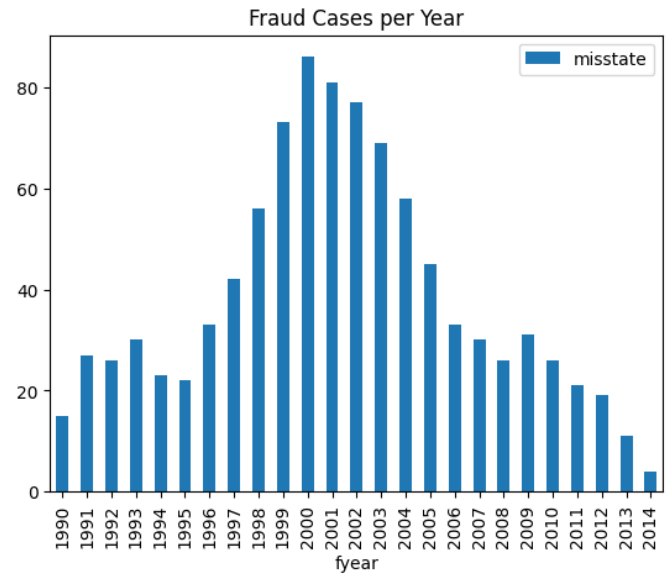


Fig. 1. Example of a figure caption.

$\text{BIB}\TeX$  does not work by magic. It doesn’t get the bibliographic data from thin air but from .bib files. If you use  $\text{BIB}\TeX$  to produce a bibliography you must send the .bib files.

$\LaTeX$  can’t read your mind. If you assign the same label to a subsection and a table, you might find that Table I has been cross referenced as Table IV-B3.

$\LaTeX$  does not have precognitive abilities. If you put a `\label` command before the command that updates the counter it’s supposed to be using, the label will pick up the last counter to be cross referenced instead. In particular, a `\label` command should not go before the caption of a figure or a table.

Do not use `\nonumber` inside the `{array}` environment. It will not stop equation numbers inside `{array}` (there won’t be any anyway) and it might stop a wanted equation number in the surrounding equation.

### E. Some Common Mistakes

- The word “data” is plural, not singular.
- The subscript for the permeability of vacuum  $\mu_0$ , and other common scientific constants, is zero with subscript formatting, not a lowercase letter “o”.
- In American English, commas, semicolons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)

- A graph within a graph is an “inset”, not an “insert”. The word alternatively is preferred to the word “alternately” (unless you really mean something that alternates).
- Do not use the word “essentially” to mean “approximately” or “effectively”.
- In your paper title, if the words “that uses” can accurately replace the word “using”, capitalize the “u”; if not, keep using lower-cased.
- Be aware of the different meanings of the homophones “affect” and “effect”, “complement” and “compliment”, “discreet” and “discrete”, “principal” and “principle”.
- Do not confuse “imply” and “infer”.
- The prefix “non” is not a word; it should be joined to the word it modifies, usually without a hyphen.
- There is no period after the “et” in the Latin abbreviation “et al.”.
- The abbreviation “i.e.” means “that is”, and the abbreviation “e.g.” means “for example”.

An excellent style manual for science writers is [?].

#### F. Authors and Affiliations

**The class file is designed for, but not limited to, six authors.** A minimum of one author is required for all conference articles. Author names should be listed starting from left to right and then moving down to the next line. This is the author sequence that will be used in future citations and by indexing services. Names should not be listed in columns nor group by affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

TABLE II  
REPEAT OFFENDERS FREQUENCY

Company Name	Frequency
Black Box Corp	13
Overseas Shipholding Group	12
Affiliated Computer Services	12
UnitedHealth Group Inc	11
Integrated Silicon Solution	9
Mercury Interactive Corp	9
Monster Worldwide Inc	8
Broadcom Corp -CI A	8
Vitesse Semiconductor Corp	8
Comverse Technology Inc	7
Elbit Medical Imaging Ltd	7
Conagra Foods Inc	7
Interpublic Group of Cos	7
NCI Inc	7
Analogic Corp	6
Dow Chemical	6
Take-Two Interactive Sftwr	6
Diebold Inc	6
NIC Inc	6
Engineered Support Systems	6
Cummins Inc	6
UCI Medical Affiliates Inc	6
Brooks Automation Inc	6
Fortune Brands Inc	6
Thor Industries Inc	6
Safenet Inc	6
Utstarcom Inc	6

#### G. Identify the Headings

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include Acknowledgments and References and, for these, the correct style to use is “Heading 5”. Use “figure caption” for your Figure captions, and “table head” for your table title. Run-in heads, such as “Abstract”, will require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced.

#### H. Figures and Tables

a) *Positioning Figures and Tables:* Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation “Fig. 2”, even at the beginning of a sentence.

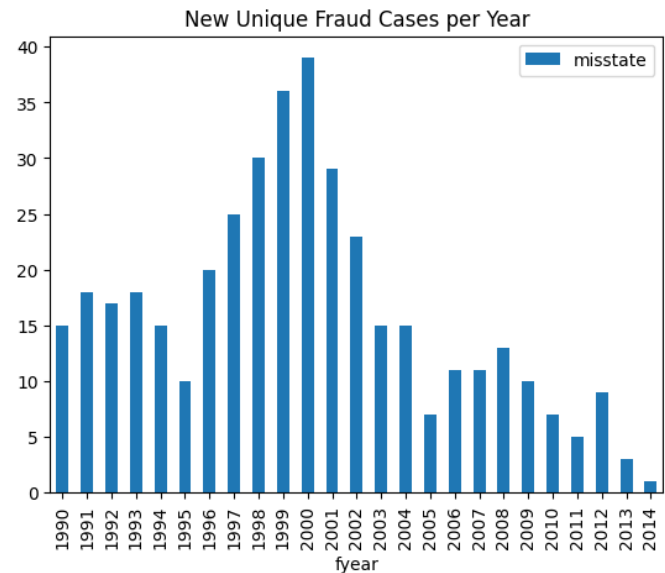


Fig. 2. Example of a figure caption.

## VI. PERFORMANCE EVALUATION METRICS

**Figure Labels:** Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an

TABLE III  
PERFORMANCE EVALUATION METRICS FOR THE TEST PERIOD 2003-08

Method	AUC	NDCG@k	Sensitivity	Precision
(1) Logit	0.703	0.022	3.103%	2.562%
(2) Random Forest	0.709	0.063	7.177%	5.714%
(3) XGBoost	0.713	0.070	8.121%	6.546%
(4) SVM	0.708	0.027	3.084%	2.285%
(5) SVM-FK	0.503	0.010	1.321%	1.130%
(6) RUSBoost	0.690	0.022	2.265%	1.997%

example, write the quantity “Magnetization”, or “Magnetization, M”, not just “M”. If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write “Magnetization (A/m)” or “Magnetization {A[m(1)]}”, not just “A/m”. Do not label axes with a ratio of quantities and units. For example, write “Temperature (K)”, not “Temperature/K”.

## VII. PRELIMINARY RESULTS

Filler

## VIII. NEXT STEPS

### A. Proposed Methods

- 1) *Transformer Architecture*: Filler
- 2) *OpenAI-Based Model*: Filler
- 3) *LLaMA (Meta)*: Filler

## ACKNOWLEDGMENT

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks ...”. Instead, try “R. B. G. thanks...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

## REFERENCES

Please number citations consecutively within brackets [?]. The sentence punctuation follows the bracket [?]. Refer simply to the reference number, as in [?]<sup>1</sup>—do not use “Ref. [?]” or “reference [?]” except at the beginning of a sentence: “Reference [?] was the first ...”

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors’ names; do not use “et al.”. Papers that have not been published, even if they have been submitted for publication, should be cited as “unpublished” [?]. Papers that have been accepted for publication should be cited as “in press” [?]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [?].

## REFERENCES

IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove the template text from your paper may result in your paper not being published.