

# Employing Machine Learning for Fraudulence Detection across Credit Card Transactions

Peep 1, 2, 3, 4, 5

CS 201R, Winter 2024

Department of Computer Science

Brigham Young University

## Abstract

Recent efforts to combat the ongoing surge in financial fraud have included the harnessment of machine learning techniques to construct anomaly detection models for various forms of payment. In this paper, we analyze the utility of machine learning algorithms in classifying the legitimacy of credit card transactions. Data was collected from (insert data sources) and subsequently pre-processed for training. We selected features from the original dataset and engineered new features to incorporate more relevant variables into the data. Novel machine learning models, including Decision Tree, Logistic Regression, and XG-Boost were implemented, trained, and evaluated against an allotted test portion of the final prepared dataset on their ability to correctly mark a credit card transaction as legitimate or fraudulent. Initial results demonstrated excellent performance ( $>99\%$  classification accuracy) with tree-based models. The high capability of these models in detecting anomalies across hundreds of thousands of transactions emphasizes the potential of machine learning for effective fraud prevention.

## 1 Introduction

Financial fraud is an evolving identity crime that affects millions worldwide annually. In particular, the advent of convenient and accessible electronic payment systems such as those linked through mobile devices, personal accounts, or software applications carries a heightened risk for financial security breaches. These emerging technologies combined with the substantial amount of transactions performed each day warrants a need for autonomous monitoring that extends beyond the practical scope of human capability. With modern advancements in data science, automatic early threat detection poses as a solution to mitigating the consequences of compromised financial information. Corporations have recently begun to implement early alert systems by training anomaly detection machine learning models on payment transaction data. Depending on the type of transaction, certain indicators may be important predictors in flagging a fraudulent payment,

such as the amount of money transacted, the time of the transaction, or whether or not a card chip was used to complete the payment.

As with many machine learning experiments, data quality, characteristics, and amount is paramount to capturing the subtle patterns and relationships necessary to make accurate predictions. Unfortunately, in our search for appropriate data, we were faced with limitations from publicly available datasets due to customer privacy concerns. Despite this bottleneck, we postulated that (with synthetic data or anonymized) general trends for financial fraud could still be replicated within the datasets and help formulate valuable and interpretable insights from prediction results. (Or if we just use anonymized data) Despite restricted interpretability in the predictions from anonymized data, realistic patterns and variations are retained. Throughout this paper, we collect and prepare transaction data, implement a variety of machine learning models, and examine classification ability through multiple metric assessments.

## 2 Methods

### 2.1 Data Collection

Print manuscripts two columns to a page, in the manner in which these instructions are printed. The exact dimensions for pages are:

- left and right margins: .75"
- column width: 3.375"
- gap between columns: .25"
- top margin—first page: 1.375"
- top margin—other pages: .75"
- bottom margin: 1.25"
- column height—first page: 6.625"
- column height—other pages: 9"

All measurements assume an  $8\frac{1}{2}'' \times 11''$  page size. For A4-size paper, use the given top and left margins, column width, height, and gap, and modify the bottom and right margins as necessary.

## 2.2 Title and Author Information

Center the title on the entire width of the page in a 14-point bold font. Below it, center the author name(s) in a 12-point bold font, and then center the address(es) in a 12-point regular font. Credit to a sponsoring agency can appear on the first page as a footnote.

## 2.3 Abstract

Place the abstract at the beginning of the first column 3" from the top of the page, unless that does not leave enough room for the title and author information. Use a slightly smaller width than in the body of the paper. Head the abstract with "Abstract" centered above the body of the abstract in a 12-point bold font. The body of the abstract should be in the same font as the body of the paper.

The abstract should be a concise, one-paragraph summary describing the general thesis and conclusion of your paper. A reader should be able to learn the purpose of the paper and the reason for its importance from the abstract. The abstract should be no more than 200 words long.

## 2.4 Text

The main body of the text immediately follows the abstract. Use 10-point type in a clear, readable font with 1-point leading (10 on 11).

Indent when starting a new paragraph, except after major headings.

## 2.5 Headings and Sections

When necessary, headings should be used to separate major sections of your paper. (These instructions use many headings to demonstrate their appearance; your paper should have fewer headings.)

### Section Headings

Print section headings in 12-point bold type in the style shown in these instructions. Leave a blank space of approximately 10 points above and 4 points below section headings. Number sections with arabic numerals.

### Subsection Headings

Print subsection headings in 11-point bold type. Leave a blank space of approximately 8 points above and 3 points below subsection headings. Number subsections with the section number and the subsection number (in arabic numerals) separated by a period.

### Subsubsection Headings

Print subsubsection headings in 10-point bold type. Leave a blank space of approximately 6 points above subsubsection headings. Do not number subsubsections.

### Special Sections

You may include an unnumbered acknowledgments section, including acknowledgments of help from colleagues, financial support, and permission to publish.

Any appendices directly follow the text and look like sections, except that they are numbered with capital letters instead of arabic numerals.

The references section is headed "References," printed in the same style as a section heading but without a number. A sample list of references is given at the end of these instructions. Use a consistent format for references, such as that provided by BibTeX. The reference list should not include unpublished work.

## 2.6 Citations

Citations within the text should include the author's last name and the year of publication, for example [?]. Append lower-case letters to the year in cases of ambiguity. Treat multiple authors as in the following examples: [?] or [?] (for more than two authors) and [?] (for two authors). If the author portion of a citation is obvious, omit it, e.g., Nebel [?]. Collapse multiple citations as follows: [?; ?].

## 2.7 Footnotes

Place footnotes at the bottom of the page in a 9-point font. Refer to them with superscript numbers.<sup>1</sup> Separate them from the text by a short line.<sup>2</sup> Avoid footnotes as much as possible; they interrupt the flow of the text.

## 3 Illustrations

Place all illustrations (figures, drawings, tables, and photographs) throughout the paper at the places where they are first discussed, rather than at the end of the paper. If placed at the bottom or top of a page, illustrations may run across both columns.

Illustrations must be rendered electronically or scanned and placed directly in your document. All illustrations should be in black and white, as color illustrations may cause problems. Line weights should be 1/2-point or thicker. Avoid screens and superimposing type on patterns as these effects may not reproduce well.

Number illustrations sequentially. Use references of the following form: Figure 1, Table 2, etc. Place illustration numbers and captions under illustrations. Leave a margin of 1/4-inch around the area covered by the illustration and caption. Use 9-point type for captions, labels, and other text in illustrations.

## Acknowledgments

The preparation of these instructions and the L<sup>A</sup>T<sub>E</sub>X and BibTeX files that implement them was supported by Schlumberger Palo Alto Research, AT&T Bell Laboratories, and Morgan Kaufmann Publishers. Preparation of the Microsoft Word file was supported by IJCAI. An early version of this document was created by Shirley Jowell and Peter F. Patel-Schneider. It was subsequently modified by Jennifer Balentine and Thomas Dean, Bernhard Nebel, and Daniel Pagenstecher. These instructions are the same as the ones for IJCAI-05, prepared by Kurt Steinkraus, Massachusetts Institute of Technology, Computer Science and Artificial Intelligence Lab.

---

<sup>1</sup>This is how your footnotes should appear.

<sup>2</sup>Note the line separating these footnotes from the text.

## A $\text{\LaTeX}$ and Word Style Files

The  $\text{\LaTeX}$  files are `ijcai09.sty` and `ijcai09.tex`, and the Bib $\text{\TeX}$  files are named `.bst` and `ijcai09.bib`. The  $\text{\LaTeX}$  style file is for version 2e of  $\text{\LaTeX}$ , and the Bib $\text{\TeX}$  style file is for version 0.99c of Bib $\text{\TeX}$  (*not* version 0.98i). The `ijcai09.sty` file is the same as the `ijcai07.sty` file used for IJCAI-07.

The Microsoft Word style file consists of a single template file, `ijcai09.dot`. This template is the same as the one used for IJCAI-07.

These Microsoft Word and  $\text{\LaTeX}$  files contain the source of the present document and may serve as a formatting sample.