FRAUD DETECTION ON BPJS INSURANCE CLAIMS USING THE ELLIPTIC ENVELOPE ALGORITHM

Neo Alit Cahya - 23522314 (*Author*)
Program Studi Magister Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jalan Ganesha 10 Bandung
E-mail (gmail): 23522314@std.stei.itb.ac.id

Abstract— **The National Health Insurance (JKN) program run by BPJS Health in Indonesia is a broad and comprehensive initiative, aimed at providing access to health for all levels of society. These programs, while valued for increasing affordability and easing access to health care, face challenges in the form of fraud that can increase costs and threaten their sustainability. Responding to potential fraud, BPJS Health has implemented an early detection strategy through the formation of a special team. This study focuses on the application of Elliptic Envelope and Support Vector Machine (SVM) methods to detect anomalies in insurance claims, identifying which is more effective in terms of scalability and accuracy. Analysis shows that Elliptic Envelope is superior in terms of processing time and evaluation scores such as silhouette_score and calinski_harabasz_score, proving its effectiveness in improving the reliability and efficiency of claims management. This research provides important insights into the application of these techniques in health insurance systems, with the goal of reducing fraud and improving resource utilization.. ***

Keywords—Fraud Detection, Elliptic Envelope, Unsupervised Learning

I. INTRODUCTION

The National Health Insurance Program (JKN) managed by BPJS Health is one of the largest health programs in the world which aims to provide broad access to health for the Indonesian population. The advantage of this program lies in its broad coverage, which allows almost all levels of society to obtain health protection. This makes JKN highly valued and trusted by the public because it simplifies access to health services and reduces the burden of individual health costs. This affordability and ease of access makes this program very popular and is key in efforts to improve public health nationally.

However, with this large scale and wide access, the JKN system becomes vulnerable to fraudulent acts which can result in cost overruns and misuse of funds. This phenomenon not only poses major financial risks but also threatens the integrity and sustainability of the existing health insurance system. Therefore, early and accurate detection of fraudulent acts is very important to save costs, prevent wastage of resources, and maintain public trust in BPJS Health management.

According to the journal "Analysis of Factors that Influence the Occurrence of BPJS Health Health Insurance Claim Fraud" (Abdullah, 2019) BPJS Health Ambon Branch once recorded a loss of Rp. 1,651,656,700,- in one year due to indications of errors made by coders in a total of 411 cases. This results in system inefficiency and potential financial deficits for BPJS Health in the future.

In the Minister of Health Regulation Number 16 of 2019, a fraud prevention team was formed at BPJS Health KC Prabumulih with the main task of preventing fraud through socialization, analysis and detection of possible fraud, followed by the necessary actions if fraud occurs (Fadliana et al., 2023) . Based on this, corrective action is needed by detecting anomalies in BPJS health insurance claim transactions to reduce company costs and increase system efficiency.

II. EASE OF USE

A. *Previous Study*

Research examining anomaly detection has been carried out by (Panda, 2024) with fraud detection in banking and (Thomas and Judith, 2020) with outlier detection in credit card transaction data. Banking transactions, credit card transactions and insurance claim transactions have several characteristics in common, namely that the data is sensitive because it contains personal data such as identity numbers, financial transactions and other personal information that must be protected to a high standard. The data is categorical and numerical, recurring transaction patterns that allow for trend and pattern analysis. Each dataset is vulnerable to fraudulent activity, whether in the form of fraudulent transactions in insurance, banking and credit card claims, whether from unauthorized transactions or identity theft.

B. Anomaly Detection in Banking

Research conducted by Panda (2024) is the application of various machine learning methods starting from Random Forest, Logistic Regression, K-Nearest Neighbors, Naive Bayes, Decision Tree, Support Vector Machine models to detect fraudulent activities in banking applications. With the significant growth of the Fintech sector, there is a need for sophisticated security systems to protect information-rich transactions from fraudulent activities. Due to the large volume of transactions and the potential for financial and reputational loss, there is an urgent need for technology that can effectively detect and prevent fraud. Several models can be used to detect outliers.
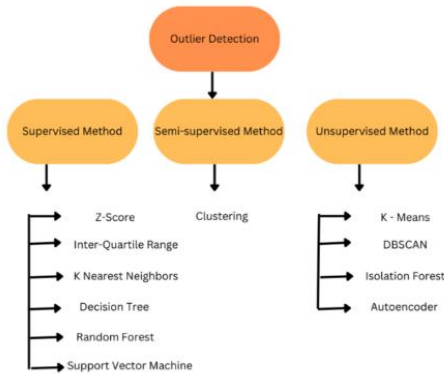

Figure 1. Outlier Detection

In research conducted by Panda (2024) used a public dataset from UCI in the form of financial transaction data with customer demographic features. To overcome imbalances in the dataset, the Synthetic Minority Over-sampling Technique (SMOTE) technique is used. Some of the methods explored in this research are Random Forest, Logistic Regression, K-Nearest Neighbors, Naive Bayes, Decision Tree, Support Vector Machine models. The proposed model successfully achieved a high accuracy of 97.74%, demonstrating its ability to identify fraudulent transactions. Disadvantages of this study include its reliance on data from a single source (UCI) which may not cover all fraud scenarios and the risk of bias from oversampling techniques.

C. Anomaly Detection in Credit Card Transactions

In research conducted by Thomas and Judith (2020) proposed an outlier detection method with unsupervised learning using an ensemble voting approach from three outlier detection algorithms, namely one-class SVM, elliptic envelope, and local outlier factor (LOF). For the iris elliptic envelope dataset, it gives the best results on the four matrices, namely 83% for Precision, 100% Recall, 90.9% F1-Score, and 0.37ms Duration

Meanwhile, for the breast cancer dataset, the ensemble model gave the best performance, especially for the 19 dimension, namely 81.5% Precision, 73.3% Recall, 77.2% F1-Score, and 0.059 ms Duration. Based on this research, the use of PCA (Principal Component Analysis) can show increased performance and efficiency, especially in high-dimensional

data. The obstacles faced in this research are that it requires appropriate parameter adjustments for each algorithm and the implementation of the model can be complex because it involves many algorithms and dimension reduction techniques. The weakness that arises in this research is that it cannot be guaranteed that the performance of the method used will be maintained if it is implemented on a larger and more diverse dataset to validate its effectiveness and scalability.

III. System Planning

PT. PLN (Persero) is a state-owned company operating in the Indonesian electricity sector. PLN has an important role in the supply and distribution of electrical energy throughout Indonesia. This company is responsible for meeting electricity needs which includes production, distribution and sales of electricity for all sectors, including industry, business and housing.

As a state company, PLN follows government regulations regarding the welfare of its employees, including health insurance. PLN employees, like other BUMN employees, are registered in the BPJS Health program. BPJS Health is a national health insurance program designed to provide health protection to all Indonesian citizens, including employees in state-owned companies.

In practice, BPJS Health insurance provides various benefits to PLN employees, ranging from preventive health protection, disease treatment, to intensive care. This is certainly very helpful in improving the quality of life of employees and their families, while ensuring that they have access to adequate and affordable health care.

The integration of health insurance systems such as BPJS Health into employee welfare packages at PLN also shows the company's commitment to fulfilling employee rights in accordance with applicable regulations. This not only benefits employees, but also helps companies maintain employee stability and performance, which ultimately impacts the company's overall efficiency and productivity.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, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

A. 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/m2" or "webers per square meter," not "webers/m2." Spell units when they appear in text: "...a few henries," not "...a few H."

- Use a zero before decimal points: "0.25," not ".25." Use "cm3," not "cc." (*bullet list*)

B. Equations

The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

Number equations consecutively. Equation numbers, within parentheses, are to position flush right, as in (1), using a right tab stop. 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

$$a + b = \gamma \tag{1}$$

Note that the equation is centered using a center tab stop. 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 ..."

C. Some Common Mistakes

- The word "data" is plural, not singular.

- The subscript for the permeability of vacuum μ_0, and other common scientific constants, is zero with subscript formatting, not a lowercase letter "o."

- In American English, commas, semi-/colons, 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 [7].

IV. USING THE TEMPLATE

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

A. Authors and Affiliations

The template is designed so that author affiliations are not repeated each time for multiple authors of the same affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization). This template was designed for two affiliations.

1) For author/s of only one affiliation (Heading 3): To change the default, adjust the template as follows.

a) Selection (Heading 4): Highlight all author and affiliation lines.

b) Change number of columns: Select the Columns icon from the MS Word Standard toolbar and then select "1 Column" from the selection palette.

c) Deletion: Delete the author and affiliation lines for the second affiliation.

2) For author/s of more than two affiliations: To change the default, adjust the template as follows.

a) Selection: Highlight all author and affiliation lines.

b) Change number of columns: Select the "Columns" icon from the MS Word Standard toolbar and then select "1 Column" from the selection palette.

c) Highlight author and affiliation lines of affiliation 1 and copy this selection.

d) Formatting: Insert one hard return immediately after the last character of the last affiliation line. Then paste down

the copy of affiliation 1. Repeat as necessary for each additional affiliation.

e) Reassign number of columns: Place your cursor to the right of the last character of the last affiliation line of an even numbered affiliation (e.g., if there are five affiliations, place your cursor at end of fourth affiliation). Drag the cursor up to highlight all of the above author and affiliation lines. Go to Column icon and select "2 Columns". If you have an odd number of affiliations, the final affiliation will be centered on the page; all previous will be in two columns.

B. 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. Styles named "Heading 1," "Heading 2," "Heading 3," and "Heading 4" are prescribed.

C. Figures and Tables

1) 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. 1," even at the beginning of a sentence.

TABLE I.   TABLE STYLES

Table Head	Table Column Head		
	Table column subhead	*Subhead*	*Subhead*
copy	More table copy[a]		

a. Sample of a Table footnote. *(Table footnote)*

b.

Fig. 1. Example of a figure caption. *(figure caption)*

We suggest that you use a text box to insert a graphic (which is ideally a 300 dpi resolution TIFF or EPS file with all fonts embedded) because this method is somewhat more stable than directly inserting a picture.

To have non-visible rules on your frame, use the MSWord "Format" pull-down menu, select Text Box > Colors and Lines to choose No Fill and No Line.

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 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."

VIDEO LINK AT YOUTUBE *(Heading 5)*

Include link of your video on YouTube in this section.

ACKNOWLEDGMENT *(Heading 5)*

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

The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use "Ref. [3]" or "reference [3]" except at the beginning of a sentence: "Reference [3] 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 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" [4]. Papers that have been accepted for publication should be cited as "in press" [5]. 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 [6].

[1] G. Eason, B. Noble, and I.N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529-551, April 1955. (*references*)

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.

[3] I.S. Jacobs and C.P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.

[4] K. Elissa, "Title of paper if known," unpublished.

[5] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740-741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

PERNYATAAN

Dengan ini saya menyatakan bahwa makalah yang saya tulis ini adalah tulisan saya sendiri, bukan saduran, atau terjemahan dari makalah orang lain, dan bukan plagiasi.

<div align="right">

Bandung, 12 Juni 2024

Ttd

Neo Alit Cahya dan 23522314

</div>