*Predicting Outcomes for Adjudication Appeals through Machine Learning*

Michaela Webster   
H. John Heinz III College  
Carnegie Mellon UniversityPittsburgh, PA, USA  
mwebster@andrew.cmu.edu

*Abstract*—Individuals who wish to work in or with the United States Government are required to undergo a background investigation. These investigations focus on investigating an individual and any and all aspects of their personal life and behavior going back 5-10 years. Some of these cases are denied and then later appealed. Of the cases that are appealed, another investigation is done and a final decision is made on whether or not the individual is granted or denied a security clearance.

Keywords—machine learning, fit, predict, model, adjudicative guidelines, scikit-learn, Jupyter Notebooks

# Introduction

This template, modified in MS Word 2007 and saved as a “Word 97-2003 Document” for the PC, provides authors with most of the formatting specifications needed for preparing electronic versions of their papers. All standard paper components have been specified for three reasons: (1) ease of use when formatting individual papers, (2) automatic compliance to electronic requirements that facilitate the concurrent or later production of electronic products, and (3) conformity of style throughout a conference proceedings. Margins, column widths, line spacing, and type styles are built-in; examples of the type styles are provided throughout this document and are identified in italic type, within parentheses, following the example. Some components, such as multi-leveled equations, graphics, and tables are not prescribed, although the various table text styles are provided. The formatter will need to create these components, incorporating the applicable criteria that follow.

# Security Clearance Process

## Guidelines

First, confirm that you have the correct template for your paper size. This template has been tailored for output on the US-letter paper size. If you are using A4-sized paper, please close this file and download the file “MSW\_A4\_format”.

## Current Process

Identify applicable funding agency here. If none, delete this text box.

The template 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 in this template 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.

# Adjudicative Appeals Process

## Guidelines

First, confirm that you have the correct template for your paper size. This template has been tailored for output on the US-letter paper size. If you are using A4-sized paper, please close this file and download the file “MSW\_A4\_format”.

## Current Process

Identify applicable funding agency here. If none, delete this text box.

The template 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 in this template 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.

# Machine Learning

Machine Learning techniques attempt to mimic human thought and make complex decisions on input using advanced statistics. There are many models that handle different forms of input and use different statistical models to predict the desired type of outcomes. In this case, Logistic Regression was used to predict outcomes due to the data input – Boolean values (true/false). Logistic Regression works by taking training data and training a classifier, creating a model from the training data, and applying that model on test data, and calculating the accuracy of the predictions against the actual values. The data that was used for the training data for this research consisted of yes/no values for various attributes and predicted whether or not the input data warranted a decision of denied or rejected.

There are also many different ways to calculate the accuracy of the model in terms of fit and how well the model has correctly predicted the value. In this case only the accuracy score was done in which the original outcomes were compared to the predicted outcomes.

# Tools

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 A-D 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 use hard tabs, and limit use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you.

## Jupyter Notebooks

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.

## Scikit-Learn

* 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 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 “cm3”, not “cc”. (*bullet list*)

# Predicting Outcomes

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.

## Scope

#### The scope of this project was focused on 5 of the 13 adjudicative guidelines. The guidelines that were used in this research were:

D: Sexual Behavior

E: Personal Conduct

F: Financial Considerations

G: Alcohol Consumption

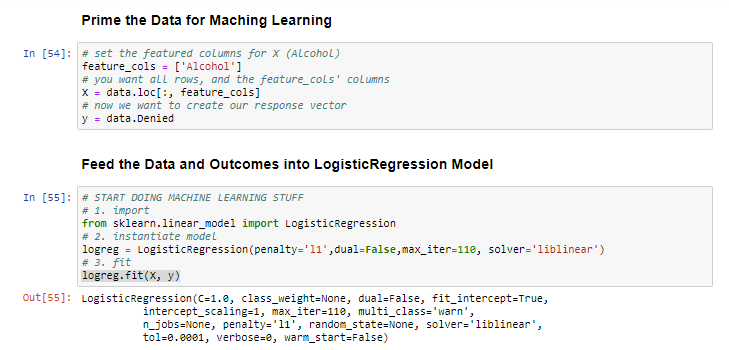
H: Drug Involvement

These guidelines were researched based on the data collected and coded and compared to the decision that was reached after the appeal process was completed (Accepted/Denied).

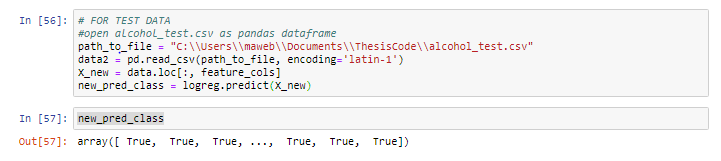
## Models Used

For these guidelines, sci-kit learn’s Logistic Regression model was used. The 3 particular functions that were used were fit, predict, and accuracy.

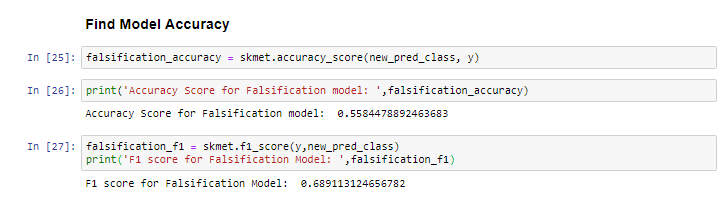
### Fit works by taking the training vector (all attributes) as X and the target vector relative to X as y. The following code snippet shows the setup for the model including assigning both X and y and the call to sci-kit learn’s fit function..



### Predict works by taking in the same parameters as X but for the test data (X\_new) in the fit scenario and outputs an array of the predicted values.



### Accuracy works by comparing the training data’s actual denied status for a case against the prediction functions denied status. F1 works by using both the precision(p) and recall(r) where p is the number of correct positive results divided by the number of all positive results returned by the classifier and r is the number of correct positive results divided by the number of all relevant samples (the samples that should have been identified as positive).



## Outcomes and Analysis

# Use Cases and Limitations

# Future Work

The models used were limited to the type of input data available. In this case, only Logistic Regression was used on a subset of input fields. Logistic Regression model does not handle text input, and since many of the other fields are not just yes/no values, branching out into more powerful models that handle more robust input is top priority on the list of things to continue to research.

Another aspect of future work is adding the most current years of data to the dataset. The current data only has information up through 2016 – adding the data from years 2017 and 2018 enhances the data points in the training data and heightens the chances for a more successful model prediction.

##### Acknowledgments

##### 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 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” [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].
7. M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.