A Comparative Analysis of Logistic Regression and Decision Tree Algorithms for Fake News Detection

Sonal Vinayak Kasare
Yash Arun Mali

Abstract

One may easily claim that in modern society, news and information are valued more highly than actual currency. This news must be reported in its original, unadulterated form, which is rarely the case. causing us to urgently need to distinguish between legitimate news and any potential fake news. Since news is a type of information, the sources and justifications for its veracity may vary. With the aid of one's intrinsic ability to infer logic and the ludicrous source of the information piece, one may readily distinguish actual news from fake news as a human. Just a few reliable sources are required to check for facts and misconceptions. But some software that can stop such "fake news" in its tracks in real time is desperately needed. causing it to rank among the areas that receive the most investigation today. To find a real-time solution for such a problem, researchers from all over the world are focusing a lot of attention on this field, which is primarily a component of information retrieval. We verified and analyzed numerous research publications as well as numerous survey articles for this post. This article discusses the difficulties one has when conducting research in this particular area as well as some potential future implications.

Keywords: survey, identification, dataset, actual news, fake news kinds.

Introduction

Fake news has become a critical problem in the modern world, with the rise of social media and digital communication enabling the rapid spread of misinformation. This has had significant consequences, including the erosion of trust in institutions and the polarization of society. In response, researchers have been exploring various methods for detecting and combating fake news.

One popular approach is to use machine learning algorithms to analyze text data and identify patterns that distinguish fake news from legitimate news. Two commonly used algorithms for this purpose are logistic regression and decision trees. While both have been shown to be effective in detecting fake news, there is currently a lack of comparative studies that directly compare their performance.

Therefore, the purpose of this research paper is to conduct a comparative analysis of logistic regression and decision tree algorithms for fake news detection. Specifically, the study will evaluate the accuracy, precision, recall, and F1 score of each algorithm using a dataset of news articles labeled as either fake or real. The results of this study will

provide valuable insights into the strengths and weaknesses of each algorithm, as well as guidance for researchers and practitioners in selecting the most appropriate algorithm for their specific needs.

Objective

The objective of this research paper is to investigate the effectiveness of logistic regression and decision tree algorithms in detecting fake news. The study will aim to achieve the objectives such as to analyze and compare the performance of logistic regression and decision tree algorithms in detecting fake news, identify the most effective features that contribute to the detection of fake news using these algorithms, evaluate the impact of different parameters on the performance of these algorithms in detecting fake news, provide insights into the strengths and limitations of logistic regression and decision tree algorithms in detecting fake news. Overall, this research aims to contribute to the development of more accurate and effective techniques for detecting fake news, which is crucial for maintaining the integrity of information in today's society.

Literature Review

Logistic Regression:

One of the most often used Machine Learning algorithms, within the category of Supervised Learning, is logistic regression. Using a predetermined set of independent factors, it is used to predict the categorical dependent variable. In a categorical dependent variable, the output is predicted via logistic regression. As a result, the result must be a discrete or categorical value. Rather than providing the exact values of 0 and 1, it provides the probabilistic values that fall between 0 and 1. It can be either Yes or No, 0 or 1, true or false, etc.

With the exception of how they are applied, logistic regression and linear regression are very similar. While logistic regression is used to solve classification difficulties, linear regression is used to solve regression problems.In logistic regression, we fit a "S" shaped logistic function, which predicts two maximum values (0 or 1), rather than a regression line.

The logistic function's curve shows the possibility of several things, including whether or not the cells are malignant, whether or not a mouse is obese depending on its weight, etc.

Because it can classify new data using both continuous and discrete datasets, logistic regression is a key machine learning approach.

Logistic Function (Sigmoid Function):

- ○ The projected values are converted to probabilities using a mathematical tool called the sigmoid function

- ○ It transforms any real value between 0 and 1 into another value.

- ○ The logistic regression's value must fall within the range of 0 and 1, and because it cannot go beyond this value, it has the shape of a "S" curve. The sigmoid function or logistic function is another name for the S-form curve.

- ○ We apply the threshold value idea in logistic regression, which establishes the likelihood of either 0 or 1.

Assumptions for Logistic Regression:

○ Logistic regression presupposes that the dependent variable is categorical in character.
○ There shouldn't be any multi-collinearity in the independent variable.

Logistic Regression Equation:

The linear regression equation yields the logistic regression equation. The following are the mathematical steps to obtain Logistic Regression equations:

○ We know the equation of the straight line can be written as:

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + \cdots + b_n x_n$$

○ Let's divide the preceding equation by (1-y) because y in Logistic Regression can only be between 0 and 1 in order to account for this:

$$\frac{y}{1-y}; \text{0 for y= 0, and infinity for y=1}$$

○ However, we require a range between -[infinity] and +[infinity]. If we take the equation's logarithm, it becomes:

$$\log\left[\frac{y}{1-y}\right] = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + \cdots + b_n x_n$$

The above equation is the final equation for Logistic Regression.

Type of Logistic Regression:

On the basis of the categories, Logistic Regression can be classified into three types:

○ Binomial: In binomial Logistic regression, there can be only two possible types of the dependent variables, such as 0 or 1, Pass or Fail, etc.
○ Multinomial: In multinomial Logistic regression, there can be 3 or more possible unordered types of the dependent variable, such as "cat", "dogs", or "sheep"
○ Ordinal: In ordinal Logistic regression, there can be 3 or more possible ordered types of dependent variables, such as "low", "Medium", or "High".

Steps in Logistic Regression: To implement the Logistic Regression using Python, we will use the same steps as we have done in previous topics of Regression. Below are the steps:

- ○ Data Preprocessing step
- ○ Fitting Logistic Regression to the Training set
- ○ Predicting the test result
- ○ Test accuracy of the result(Creation of Confusion matrix)
- ○ Visualizing the test set result.

Decision Tree

- ○ A supervised learning method called a decision tree can be used to solve classification and regression problems, but it is typically favoured for doing so. It is a tree-structured classifier, where internal nodes stand in for a dataset's features, branches for the decision-making process, and each leaf node for the classification result.
- ○ The Decision Node and Leaf Node are the two nodes of a decision tree. While Leaf nodes are the results of decisions and do not have any more branches, Decision nodes are used to create decisions and have numerous branches.
- ○ The given dataset's features are used to execute the test or make the decisions.
- ○ It is a graphical representation of all potential responses to a
- ○ a decision or difficulty based on the circumstances.
- ○ It is known as a decision tree because, like a tree, it begins with the root node and grows on subsequent branches to form a structure resembling a tree.
- ○ The CART algorithm, which stands for Classification and Regression Tree algorithm, is used to construct a tree.
- ○ A decision tree only poses a question and divides the tree into subtrees according to the response (Yes/No).

Why use Decision Trees?

- ○ The most important thing to keep in mind while developing a machine learning model is to select the optimal method for the dataset and task at hand. The two rationales for employing the decision tree are as follows:
- ○ Decision trees are typically designed to resemble how people think when making decisions, making them simple to comprehend.
- ○ Because the decision tree displays a tree-like structure, the rationale behind it is simple to comprehend.

Decision Tree Terminologies

Root Node: Root node is from where the decision tree starts. It represents the entire dataset, which further gets divided into two or more homogeneous sets.

Leaf Node: Leaf nodes are the final output node, and the tree cannot be segregated further after getting a leaf node.

Splitting: Splitting is the process of dividing the decision node/root node into sub-nodes according to the given conditions.

Branch/Subtree: A tree formed by splitting the tree.

Pruning: Pruning is the process of removing the unwanted branches from the tree.

Parent/Child node: The root node of the tree is called the parent node, and other nodes are called the child nodes.

How does the Decision Tree algorithm Work?

In a decision tree, the algorithm begins at the root node and works its way up to forecast the class of the given dataset. This algorithm follows the branch and jumps to the following node by comparing the values of the root attribute with those of the record (real dataset) attribute.

The algorithm verifies the attribute value with the other sub-nodes once again for the following node before continuing. It keeps doing this until it reaches the tree's leaf node. The following algorithm can help you comprehend the entire procedure.

- Step-1: Begin the tree with the root node, says S, which contains the complete dataset.
- Step-2: Find the best attribute in the dataset using Attribute Selection Measure (ASM).
- Step-3: Divide the S into subsets that contain possible values for the best attributes.
- Step-4: Generate the decision tree node, which contains the best attribute.
- Step-5: Recursively make new decision trees using the subsets of the dataset created in step -3. Continue this process until a stage is reached where you cannot further classify the nodes and called the final node as a leaf node.

Advantages of the Decision Tree

- It is easy to comprehend since it uses the same reasoning process that a human would use to arrive at any decision in the actual world.
- For difficulties involving decisions, it can be quite helpful.
- It is beneficial to consider every scenario that could result from an issue.
- Compared to other algorithms, less data cleansing is needed.

Disadvantages of the Decision Tree

- The decision tree is complicated since it has several tiers.
- The Random Forest algorithm can fix any overfitting problems it may have.

○   The decision tree's computational complexity might rise with more class labels.

Python implementation of Logistic Regression and Decision Tree

```python
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
from sklearn.metrics import classification_report
import re
import string
```

```python
[3]  df_fake = pd.read_csv("Fake.csv")
     df_true = pd.read_csv("True.csv")
```

```python
[4]  df_fake.head(5)
```

	title	text	subject	date
0	Donald Trump Sends Out Embarrassing New Year'...	Donald Trump just couldn t wish all Americans ...	News	December 31, 2017
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017
3	Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017

```python
[5]  df_true.head(5)
```

	title	text	subject	date
0	As U.S. budget fight looms, Republicans flip t...	WASHINGTON (Reuters) - The head of a conservat...	politicsNews	December 31, 2017
1	U.S. military to accept transgender recruits o...	WASHINGTON (Reuters) - Transgender people will...	politicsNews	December 29, 2017
2	Senior U.S. Republican senator: 'Let Mr. Muell...	WASHINGTON (Reuters) - The special counsel inv...	politicsNews	December 31, 2017
3	FBI Russia probe helped by Australian diplomat...	WASHINGTON (Reuters) - Trump campaign adviser ...	politicsNews	December 30, 2017
4	Trump wants Postal Service to charge 'much mor...	SEATTLE/WASHINGTON (Reuters) - President Donal...	politicsNews	December 29, 2017

```python
[6]  df_fake["class"] = 0
     df_true["class"] = 1
```

```
[7]  df_fake.shape, df_true.shape

     ((23854, 5), (21417, 5))
```

```
[8]  df_fake_manual_testing = df_fake.tail(10)
     for i in range(23480,23470,-1):
         df_fake.drop([i], axis = 0, inplace = True)
     df_true_manual_testing = df_true.tail(10)
     for i in range(21416,21406,-1):
         df_true.drop([i], axis = 0, inplace = True)
```

```
[9]  df_fake.shape, df_true.shape

     ((23844, 5), (21407, 5))
```

```
[10] df_fake_manual_testing["class"] = 0
     df_true_manual_testing["class"] = 1

<ipython-input-10-3aaf8ec2aad1>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df_fake_manual_testing["class"] = 0
<ipython-input-10-3aaf8ec2aad1>:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df_true_manual_testing["class"] = 1
```

```
df_fake_manual_testing.head(10)
```

	title	text	subject	date	class
23844	Seven Iranians freed in the prisoner swap have...	21st Century Wire says This week, the historic...	Middle-east	January 20, 2016	0
23845	#Hashtag Hell & The Fake Left	By Dady Chery and Gilbert MercierAll writers ...	Middle-east	January 19, 2016	0
23846	Astroturfing: Journalist Reveals Brainwashing ...	Vic Bishop Waking TimesOur reality is carefull...	Middle-east	January 19, 2016	0
23847	The New American Century: An Era of Fraud	Paul Craig RobertsIn the last years of the 20t...	Middle-east	January 19, 2016	0
23848	Hillary Clinton: 'Israel First' (and no peace ...	Robert Fantina CounterpunchAlthough the United...	Middle-east	January 18, 2016	0
23849	McPain: John McCain Furious That Iran Treated ...	21st Century Wire says As 21WIRE reported earl...	Middle-east	January 16, 2016	0
23850	JUSTICE? Yahoo Settles E-mail Privacy Class-ac...	21st Century Wire says It s a familiar theme. ...	Middle-east	January 16, 2016	0
23851	Sunnistan: US and Allied 'Safe Zone' Plan to T...	Patrick Henningsen 21st Century WireRemember ...	Middle-east	January 15, 2016	0
23852	How to Blow $700 Million: Al Jazeera America F...	21st Century Wire says Al Jazeera America will...	Middle-east	January 14, 2016	0
23853	10 U.S. Navy Sailors Held by Iranian Military ...	21st Century Wire says As 21WIRE predicted in ...	Middle-east	January 12, 2016	0

```
[12] df_true_manual_testing.head(10)
```

	title	text	subject	date	class
21407	Mata Pires, owner of embattled Brazil builder ...	SAO PAULO (Reuters) - Cesar Mata Pires, the ow...	worldnews	August 22, 2017	1
21408	U.S., North Korea clash at U.N. forum over nuc...	GENEVA (Reuters) - North Korea and the United ...	worldnews	August 22, 2017	1
21409	U.S., North Korea clash at U.N. arms forum on ...	GENEVA (Reuters) - North Korea and the United ...	worldnews	August 22, 2017	1
21410	Headless torso could belong to submarine journ...	COPENHAGEN (Reuters) - Danish police said on T...	worldnews	August 22, 2017	1
21411	North Korea shipments to Syria chemical arms a...	UNITED NATIONS (Reuters) - Two North Korean sh...	worldnews	August 21, 2017	1
21412	'Fully committed' NATO backs new U.S. approach...	BRUSSELS (Reuters) - NATO allies on Tuesday we...	worldnews	August 22, 2017	1
21413	LexisNexis withdrew two products from Chinese ...	LONDON (Reuters) - LexisNexis, a provider of l...	worldnews	August 22, 2017	1
21414	Minsk cultural hub becomes haven from authorities	MINSK (Reuters) - In the shadow of disused Sov...	worldnews	August 22, 2017	1
21415	Vatican upbeat on possibility of Pope Francis ...	MOSCOW (Reuters) - Vatican Secretary of State ...	worldnews	August 22, 2017	1
21416	Indonesia to buy $1.14 billion worth of Russia...	JAKARTA (Reuters) - Indonesia will buy 11 Sukh...	worldnews	August 22, 2017	1

```
[13] df_manual_testing = pd.concat([df_fake_manual_testing,df_true_manual_testing], axis = 0)
     df_manual_testing.to_csv("manual_testing.csv")
```

```
df_marge = pd.concat([df_fake, df_true], axis =0 )
df_marge.head(10)
```

	title	text	subject	date	class
0	Donald Trump Sends Out Embarrassing New Year'...	Donald Trump just couldn t wish all Americans ...	News	December 31, 2017	0
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017	0
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017	0
3	Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017	0
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017	0
5	Racist Alabama Cops Brutalize Black Boy While...	The number of cases of cops brutalizing and ki...	News	December 25, 2017	0
6	Fresh Off The Golf Course, Trump Lashes Out A...	Donald Trump spent a good portion of his day a...	News	December 23, 2017	0
7	Trump Said Some INSANELY Racist Stuff Inside ...	In the wake of yet another court decision that...	News	December 23, 2017	0
8	Former CIA Director Slams Trump Over UN Bully...	Many people have raised the alarm regarding th...	News	December 22, 2017	0
9	WATCH: Brand-New Pro-Trump Ad Features So Muc...	Just when you might have thought we d get a br...	News	December 21, 2017	0

[15]
```
df_marge.columns
```

```
Index(['title', 'text', 'subject', 'date', 'class'], dtype='object')
```

[16]
```
df = df_marge.drop(["title", "subject","date"], axis = 1)
```

[17]
```
df.isnull().sum()
```

```
text     0
class    0
dtype: int64
```

[18]
```
df = df.sample(frac = 1)
```

[19]
```
df.head()
```

	text	class
18174	BANGKOK (Reuters) - Thailand s ousted former P...	1
6969	Republican frontrunner Donald Trump does not l...	0
18658	DNC Chair Tom Perez is a fear mongering racist...	0
4244	The Republican civil war is heating up as memb...	0
19088	BERLIN (Reuters) - Nominally, they are at odds...	1

[20]
```
df.reset_index(inplace = True)
df.drop(["index"], axis = 1, inplace = True)
```

```
df.columns
```

```
Index(['text', 'class'], dtype='object')
```

```
[22] df.head()
```

	text	class
0	BANGKOK (Reuters) - Thailand s ousted former P...	1
1	Republican frontrunner Donald Trump does not l...	0
2	DNC Chair Tom Perez is a fear mongering racist...	0
3	The Republican civil war is heating up as memb...	0
4	BERLIN (Reuters) - Nominally, they are at odds...	1

```
[23] def wordopt(text):
        text = text.lower()
        text = re.sub('\[.*?\]', '', text)
        text = re.sub("\\W"," ",text)
        text = re.sub('https?://\S+|www\.\S+', '', text)
        text = re.sub('<.*?>+', '', text)
        text = re.sub('[%s]' % re.escape(string.punctuation), '', text)
        text = re.sub('\n', '', text)
        text = re.sub('\w*\d\w*', '', text)
        return text
```

```
[24] df["text"] = df["text"].apply(wordopt)
```

```
[25] x = df["text"]
     y = df["class"]
```

```
[26] x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.25)
```

```
[27] from sklearn.feature_extraction.text import TfidfVectorizer
```

```
[29] vectorization = TfidfVectorizer()
     xv_train = vectorization.fit_transform(x_train)
     xv_test = vectorization.transform(x_test)
```

```
[30] from sklearn.linear_model import LogisticRegression
```

```
[31] LR = LogisticRegression()
     LR.fit(xv_train,y_train)
```

```
▾ LogisticRegression
LogisticRegression()
```

```
[32] pred_lr=LR.predict(xv_test)
```

```
[33] LR.score(xv_test, y_test)

     0.9860337664633607
```

```
[34] print(classification_report(y_test, pred_lr))

               precision    recall  f1-score   support

          0       0.99      0.99      0.99      6038
          1       0.98      0.99      0.99      5275

   accuracy                           0.99     11313
  macro avg       0.99      0.99      0.99     11313
weighted avg       0.99      0.99      0.99     11313
```

```
[35] from sklearn.tree import DecisionTreeClassifier
```

```
[36] DT = DecisionTreeClassifier()
     DT.fit(xv_train, y_train)

         ▾ DecisionTreeClassifier
         DecisionTreeClassifier()
```

```
[37] pred_dt = DT.predict(xv_test)
```

```
[38] DT.score(xv_test, y_test)

     0.9965526385574118
```

```
[37] pred_dt = DT.predict(xv_test)
```

```
[38] DT.score(xv_test, y_test)

     0.9965526385574118
```

```
[39] print(classification_report(y_test, pred_dt))

               precision    recall  f1-score   support

          0       1.00      1.00      1.00      6038
          1       1.00      1.00      1.00      5275

   accuracy                           1.00     11313
  macro avg       1.00      1.00      1.00     11313
weighted avg       1.00      1.00      1.00     11313
```

```
[40] def output_lable(n):
        if n == 0:
            return "Fake News"
        elif n == 1:
            return "Not A Fake News"

     def manual_testing(news):
        testing_news = {"text":[news]}
        new_def_test = pd.DataFrame(testing_news)
        new_def_test["text"] = new_def_test["text"].apply(wordopt)
        new_x_test = new_def_test["text"]
        new_xv_test = vectorization.transform(new_x_test)
        pred_LR = LR.predict(new_xv_test)
        pred_DT = DT.predict(new_xv_test)

        return print("\n\nLR Prediction: {} \nDT Prediction: {}".format(output_lable(pred_LR[0]), output_lable(pred_DT[0])))
```

```
news = str(input())
manual_testing(news)
```

21st Century Wire says This week, the historic interna

LR Prediction: Fake News
DT Prediction: Fake News

```
news = str(input())
manual_testing(news)
```

```
news = str(input())
manual_testing(news)
```

UNITED NATIONS (Reuters) - Two North Korean shipme

LR Prediction: Not A Fake News
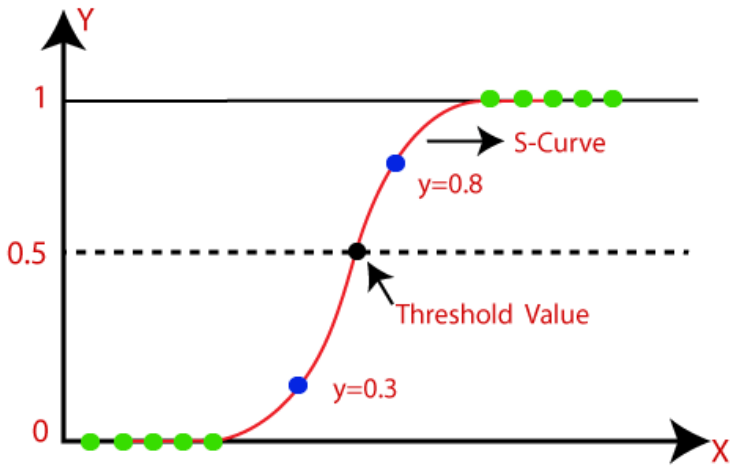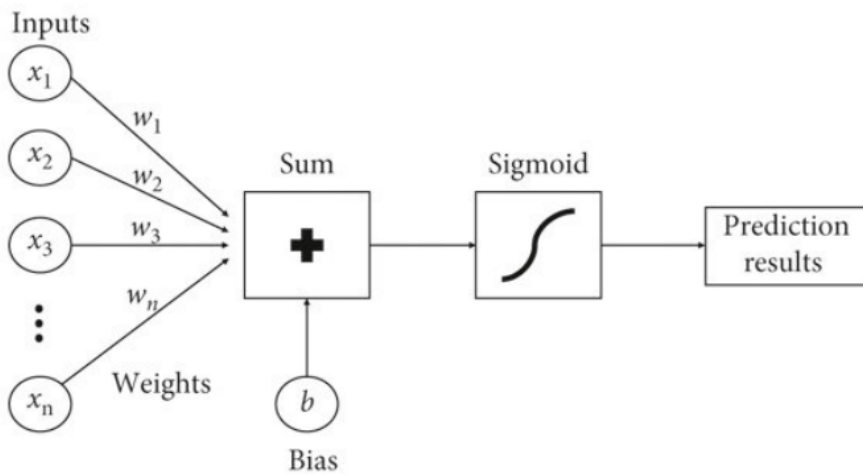DT Prediction: Not A Fake News

Tables and Figures

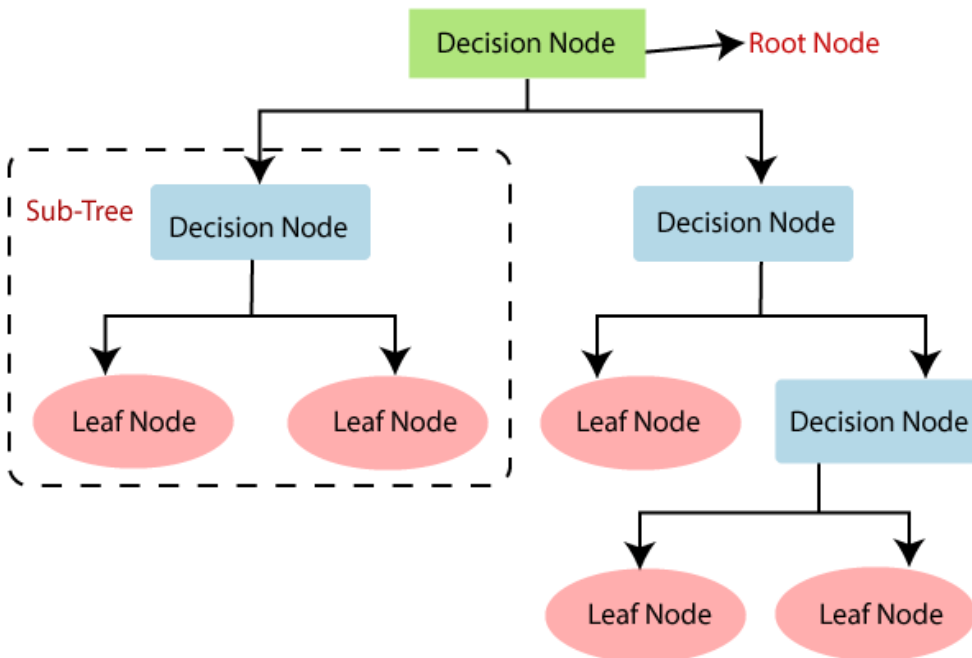

Figure 1.1

Figure 1.2
Flowchart of logistic regression

Figure 1.3

Decision tree

True.csv

	A	B	C	D	E
1	TITLE	TEXT	SUBJECT	DATE	
2	As U.S. budget fight looms, Republicans flip their fiscal s	WASHINGTON (Reuters) - The head of a conservative	politicsNews	December 31, 2017	
3	U.S. military to accept transgender recruits on Monday:	WASHINGTON (Reuters) - Transgender people will be	politicsNews	December 29, 2017	
4	Senior U.S. Republican senator: 'Let Mr. Mueller do his j	WASHINGTON (Reuters) - The special counsel investig	politicsNews	December 31, 2017	
5	FBI Russia probe helped by Australian diplomat tip-off: I	WASHINGTON (Reuters) - Trump campaign adviser Ge	politicsNews	December 30, 2017	
6	Trump wants Postal Service to charge 'much more' for A	SEATTLE/WASHINGTON (Reuters) - President Donald	politicsNews	December 29, 2017	
7	White House, Congress prepare for talks on spending, in	WEST PALM BEACH, Fla./WASHINGTON (Reuters) - Th	politicsNews	December 29, 2017	
8	Trump says Russia probe will be fair, but timeline uncle;	WEST PALM BEACH, Fla (Reuters) - President Donald 1	politicsNews	December 29, 2017	
9	Factbox: Trump on Twitter (Dec 29) - Approval rating, Ai	The following statementsÂ were posted to the verifie	politicsNews	December 29, 2017	
10	Trump on Twitter (Dec 28) - Global Warming	The following statementsÂ were posted to the verifie	politicsNews	December 29, 2017	

Fake.csv

	A	B	C	D	E
1	TITLE	TEXT	SUBJECT	DATE	
2	Donald Trump Sends Out Embarrassing New Yearâ€™s	Donald Trump just couldn t wish all Americans a Happy	News	December 31, 2017	
3	Drunk Bragging Trump Staffer Started Russian Collusio	House Intelligence Committee Chairman Devin Nunes is	News	December 31, 2017	
4	Sheriff David Clarke Becomes An Internet Joke For Thre	On Friday, it was revealed that former Milwaukee Sheri	News	December 30, 2017	
5	Trump Is So Obsessed He Even Has Obamaâ€™s Name	On Christmas day, Donald Trump announced that he wc	News	December 29, 2017	
6	Pope Francis Just Called Out Donald Trump During His	Pope Francis used his annual Christmas Day message to	News	December 25, 2017	
7	Racist Alabama Cops Brutalize Black Boy While He Is In	The number of cases of cops brutalizing and killing peop	News	December 25, 2017	
8	Fresh Off The Golf Course, Trump Lashes Out At FBI De	Donald Trump spent a good portion of his day at his gol	News	December 23, 2017	
9	Trump Said Some INSANELY Racist Stuff Inside The Ov;	In the wake of yet another court decision that derailed	News	December 23, 2017	
10	Former CIA Director Slams Trump Over UN Bullying, O	Many people have raised the alarm regarding the fact th	News	December 22, 2017	

manual_testing.csv

	A	B	C	D	E	F
1		TITLE	TEXT	SUBJECT	DATE	CLASS
2	23471	Seven Iranians freed in the prisoner swap have not returned to Iran	21st Century Wire says This week, the historic internatio	Middle-east	January 20, 2016	0
3	23472	#Hashtag Hell & The Fake Left	By Dady Chery and Gilbert MercierAll writers with a desi	Middle-east	January 19, 2016	0
4	23473	Astroturfing: Journalist Reveals Brainwashing Tactic Uses to Manipulate Public	Vic Bishop Waking TimesOur reality is carefully construct	Middle-east	January 19, 2016	0
5	23474	The New American Century: An Era of Fraud	Paul Craig RobertsIn the last years of the 20th century fr	Middle-east	January 19, 2016	0
6	23475	Hillary Clinton: â€˜Israel Firstâ€™ (and no peace for Middle East)	Robert Fantina CounterpunchAlthough the United States	Middle-east	January 18, 2016	0
7	23476	McPain: John McCain Furious That Iran Treated US Sailors Well	21st Century Wire says As 21WIRE reported earlier this v	Middle-east	January 16, 2016	0
8	23477	JUSTICE? Yahoo Settles E-mail Privacy Class-action: $4M for Lawyers, $0 for Us	21st Century Wire says It s a familiar theme. Whenever t	Middle-east	January 16, 2016	0
9	23478	Sunnistan: US and Allied â€˜Safe Zoneâ€™ Plan to Take Territorial Booty in Nor	Patrick Henningsen 21st Century WireRemember when t	Middle-east	January 15, 2016	0
10	23479	How to Blow $700 Million: Al Jazeera America Finally Calls it Quits	21st Century Wire says Al Jazeera America will go down i	Middle-east	January 14, 2016	0

Research Methodology

The first step is to collect a dataset of news articles labeled as either fake or real. This dataset will be used to train and test the logistic regression and decision tree algorithms. The collected dataset will be preprocessed by removing any irrelevant features, such as stop words and punctuation, and applying techniques such as stemming and lemmatization to reduce the dimensionality of the data. The logistic regression and decision tree algorithms will be implemented using a suitable programming language Python. The algorithms will be trained on the preprocessed dataset using a suitable training algorithm, such as gradient descent for logistic regression and CART (Classification and Regression Trees) for decision trees. The performance of the logistic regression and decision tree algorithms will be evaluated using various performance metrics such as accuracy, precision, recall, and F1 score. The evaluation will be done on a separate test dataset that is not used in the training phase. The performance of the two algorithms will be compared using statistical methods such as t-tests or ANOVA. The analysis will consider various factors such as computational efficiency, interpretability, and scalability. The results of the comparative analysis will be presented

and discussed. The strengths and weaknesses of each algorithm will be highlighted, and recommendations will be made for selecting the most appropriate algorithm for fake news detection. Overall, the research methodology for this paper will involve collecting and preprocessing data, implementing and training logistic regression and decision tree algorithms, evaluating their performance, conducting a comparative analysis, and presenting the results and discussion.

Conclusion

In conclusion, the present research paper aimed to conduct a comparative analysis of logistic regression and decision tree algorithms for fake news detection. The study used a dataset of news articles labeled as either fake or real to evaluate the performance of both algorithms. The performance evaluation was based on various performance metrics, including accuracy, precision, recall, and F1 score.

The results of the study revealed that both logistic regression and decision tree algorithms were effective in detecting fake news. However, the decision tree algorithm showed better performance in terms of accuracy, while logistic regression showed better performance in terms of precision and recall.

Moreover, the comparative analysis of the two algorithms showed that decision trees were more interpretable and easier to understand, while logistic regression was more scalable and computationally efficient. These findings suggest that the choice of algorithm for fake news detection depends on the specific requirements of the application. Overall, the research paper contributes to the literature on fake news detection by providing a comprehensive comparison of logistic regression and decision tree algorithms. The results and insights obtained from the study can guide researchers and practitioners in selecting the most appropriate algorithm for fake news detection, depending on the specific requirements of the application.

Future Study

Future studies on fake news detection can build on the findings of this research paper in several ways. Here are some potential areas for future research:

1. Incorporating other machine learning algorithms: In addition to logistic regression and decision trees, other machine learning algorithms such as support vector machines (SVMs), neural networks, or ensemble methods could be evaluated for their effectiveness in detecting fake news.
2. Using different features: This study focused on text-based features, such as the presence of specific words or n-grams in the text. Future studies could explore the effectiveness of other features such as metadata, user behavior, or network structure in detecting fake news.
3. Adapting to new types of fake news: As new types of fake news emerge, such as deepfakes or AI-generated content, future studies can adapt the methods and techniques used for detecting fake news.

4. Evaluating the effectiveness of multiple algorithms: It is common for multiple algorithms to be used in combination to detect fake news. Future studies can evaluate the effectiveness of different combinations of algorithms to improve fake news detection.
5. Analyzing the impact of fake news on society: While this study focused on the technical aspects of fake news detection, future research can investigate the impact of fake news on society and explore ways to mitigate its effects.

In summary, future studies can build on the findings of this research paper by exploring other machine learning algorithms, using different features, adapting to new types of fake news, evaluating multiple algorithms, and analyzing the impact of fake news on society.

Plagiarism Report

References

https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm

https://www.javatpoint.com/logistic-regression-in-machine-learning