|  |
| --- |
| **TITLE PAGE**  **FAKE NEWS DETECTOR, A CHROME EXTENSION**  **By**  **BIBEK GAUTAM**  **Dissertation Submitted to the engineering science and Technology, Infrastructure University Kuala Lumpur, in Fulfillment of the Requirements for the Bachelor of Computer Science (Hons)**  **2021** |

# **ABSTRACT**

In the present world, millions of information have been consumed every day from different sources, among which most of them are not authenticated. The existing extensions verify whether the selected article is fake or real but does not provide any cross-reference. To overcome such a problem, *Fake news detector, A chrome extension* is designed which is a browser extension for Google chrome. It identifies news credibility with supporting cross-reference. Authentication requires numerous manpower and consumes much time to perform the task. For this, an ML model is introduced to identify the authenticity of the news article, and to support its decision, references on a similar article are provided through different trusted sites. To clearly understand the system and its flow, agile methodology with unified modeling language design is practiced. This project aims to develop a chrome extension that is reliable and easy to detect the authentication of the article. The proposed system takes input from the users, processes that article through the Machine learning module then displays the output with the list of trusted sites if that particular article is valid. A simple user interface is applied to the system to enhance the user’s performance while surfing through the internet. The limitation of this system is tested through various testing methods, the results of the project are mentioned. Overall the project is concluded at the end and the future work is highlighted.

**ACKNOWLEDGEMENT**

Luckily for all the things all through my way of information exploration to finish this project.

I might want to offer my earnest thanks to my supervisor and furthermore my teacher Mr. Akash Deo for his direction and guidance all through this work to make it fruitful.

My most profound appreciation to my family particularly my folks for their most extreme help and support without them all these would not be conceivable. To wrap things up, an abundance of thanks and appreciation to every one of my companions for assisting me with a ton of thoughts, inspirations, and supportive gestures

**APPROVAL**

We have inspected this exposition and checked that it meets the program and school necessities for the Bachelor of Computer Science BCS (Hons.)

Signature: ..........................

Supervisor: Mr. Akash Deo

Date: 31st August 2021

**AGREEMENT LETTER**

**Legitimate Reasons for Late Submission**

* Students are needed to present their Chapter Assignments to their Supervisor.
* Students ought to guarantee they know about how and when to present their tasks.
* Students who miss the deadlines on account of illness should introduce an original (not a copy) clinical testament to the instructor on the first day back in quite a while.
* Students who solicit sympathetic thought should give them a memorial service declaration, eulogy, specialist's authentication or demise testament, or pertinent documentation.

**Students Declaration**

* I announce that this project is my own work and doesn't include plagiarism or conspiracy and will be reviewed as Failed whenever demonstrated.
* It doesn't contain my own past work without this being expressed.
* I'm mindful that any violation of these principles will be viewed as cheating and considered as Failed.
* I'm mindful that all material submitted will be acknowledged after the deadline yet no marks will be given.
* I have made a copy and electronic copy of my task, which I can create if the first is lost in any way, shape, or form.
* This declaration will be naturally removed once I drop the IT Project subject.

Student Name: Bibek Gautam Student Signature: \_\_\_\_\_\_\_\_\_\_\_\_\_\_

Student ID: 041802900011 Date: 31/08/2021

**DECLARATION**

I affirm that the project report and the research to which it alludes are the consequences of my work and that any recommendations or statements from others' work, composed or something else, are totally perceived inconsistency with the discipline's standard reference practices.

August 31, 2021 Bibek Gautam

041802900011

**TABLE OF CONTENTS**

**TITLE PAGE i**

**ABSTRACT ii**

**ACKNOWLEDGEMENT iii**

**APPROVAL iv**

**AGREEMENT LETTER v**

**DECLARATION vi**

**TABLE OF CONTENTS vii**

**LIST OF FIGURES ix**

**LIST OF TABLE x**

**LIST OF ABBREVIATIONS x**

[**CHAPTER-1**](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.isbyvm85leed)[**1**](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.isbyvm85leed)

[**INTRODUCTION**](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.epgkrvsv9sl0)[**1**](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.epgkrvsv9sl0)

1.1[Introduction](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.detik3h550gh) [1](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.detik3h550gh)

1.2 [Problem Statement](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.o8hbohjk7rde) [2](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.o8hbohjk7rde)

1.3 [Objective](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.3yfrgugaav6l) [2](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.3yfrgugaav6l)

1.4 [Scope](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.j21wts20vduj) 2

1.4.1 [User Scope](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.684u2lgnx9z3) 2

1.4.2 [System Scope](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.ptp5lu2980bm) 2

1.4.3 [Constraints](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.n16nf6ew1z1) [3](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.n16nf6ew1z1)

1.5 [Methodology](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.nska5x6jgd6w) 4

1.6 Development tools 4

1.6.1 Machine Learning 4

1.6.2 Javascript 4

1.6.3 Python 4

1.6.4 VS code 5

1.6.5 Draw.io 5

1.7 [Conclusion](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.55sv7p987wiu) 5

[**CHAPTER - 2**](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.a0sre9e7kqyg)[**7**](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.a0sre9e7kqyg)

[**LITERATURE REVIEW**](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.wngrs2s25a12)[**7**](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.wngrs2s25a12)

2.1 Introduction 6

2.2[Background study](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.lwdhh2b9i326) 7

2.3 [Algorithm](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.lv2rnc5mtfed) [1](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.lv2rnc5mtfed)0

2.3.1 Logistic Regression 10

2.3.2 Support Vector Machine 10

2.3.3 Multilayer Perceptron 10

2.3.4 Passive-Aggressive classifiers 10

2.3.5 Decision Tree classifier 11

2.3.6 Random Forest Classifier 11

2.4 Common feature of the similar system 11

2.4.1 Text resemblance 11

2.4.2 Domain type detection 11

2.4.3 Deep syntax analysis 11

2.4.4 Fact-checking 12

2.4.5 Predictive modeling-based methods 12

2.4.6 Text Classification 12

2.5 [Research of similar systems](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.wssfwkvloy7r) [1](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.wssfwkvloy7r)2

2.5.1 Fake News Detector 12

2.5.2 FakerFact: Fake News Detection 13

2.5.3 Bullshit Detector 14

2.5.4 Fake News Detector with Pinocchio Alerts 14

2.5.5 Debunker - Detect Fake News 15

2.5.6 No Bias News 16

2.5.7 120 Lies: Fact-Check Amalgamator 16

2.5.8 Centr: Political Bias Detector 17

2.5.9 NewsCracker 18

2.5.10 FakeNews Checker 19

2.6 Comparison Table 20

2.7 Conclusion 20

[**CHAPTER - 3**](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.gwv7hv70bkb3)[**2**](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.gwv7hv70bkb3)**1**

[**METHODOLOGY**](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.dvz6rn5zc7b2)[**2**](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.dvz6rn5zc7b2)**1**

3.1 [Introduction](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.wfrd7ljbbr3g) [2](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.wfrd7ljbbr3g)1

3.2 [Agile Model](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.vzyrjgnvjjmb) [2](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.vzyrjgnvjjmb)1

3.3 Use case Diagram 24

3.4 [Written Use Case](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.cd7loa86emqx) [2](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.cd7loa86emqx)5

3.4.1 Select News 25

3.4.2 Chrome Extension Enabled 25

3.4.3 ML Module 26

3.4.4 Cross-Referencing 26

3.4.5 Display Result 27

3.5 Activity Diagram 28

3.6 [Sequence Diagram](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.dcfwbppekcde) [2](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.dcfwbppekcde)9

3.7 [Class Diagram](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.6yzhruczrb1y) 29

3.8 [Conclusion](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.lm3jah2uk7w6) [3](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.lm3jah2uk7w6)0

[**CHAPTER 4**](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.no3nwf0uji)[**3**](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.no3nwf0uji)**1**

[**INTERFACE**](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.u1uqf73cpqsq)[**3**](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.u1uqf73cpqsq)**1**

4.1[Introduction](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.t4kwwde5b81k) [3](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.t4kwwde5b81k)1

4.2 [Interface explain](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.y1052rc3zavr) [3](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.y1052rc3zavr)1

4.2.1 [Extension Tab](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.6ojs3ifza93j) [3](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.6ojs3ifza93j)1

4.2.2 [Main Modal Window](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.8fnp6des96qz) [3](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.8fnp6des96qz)2

4.2.3 [Result Modal Window (Real News)](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.1m0zz0tm53cm) [3](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.1m0zz0tm53cm)2

4.2.4 [Result Modal Window (Fake News)](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.9tpezq7xf1y8) [3](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.9tpezq7xf1y8)3

4.3 [Conclusion](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.ksnih56pczi5) [3](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.ksnih56pczi5)4

[**CHAPTER 5**](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.glompev2tb2r)[**3**](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.glompev2tb2r)**5**

[**TESTING AND RESULT**](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.7c54gsh8ym7e)[**3**](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.7c54gsh8ym7e)**5**

5.1 [Introduction](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.suegzn3eg1as) [3](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.suegzn3eg1as)5

5.2 [Dataset](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.saomihaueuzn) [3](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.saomihaueuzn)6

5.3 [ML Model Results](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.rg1c5ianjjcd) 37

5.3.1 [Logistic Regression Model](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.c89z4g39n7) 37

5.3.2 [MultinomiaLNB (naive\_bayes) Model](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.f1zoto7prgm) 38

5.3.3 [Decision Tree Classifier Model](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.49pg2w1r2uup) 38

5.3.4 [PassiveAggressive Classifier Model](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.uqnmvi3anhag) 39

5.3.5 [Random Forest Classifier Model](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.1hqu6kmekfjx) [4](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.1hqu6kmekfjx)0

5.3.6 [LinearSVC Model](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.nqfkb57xy5ln) [4](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.nqfkb57xy5ln)0

5.4 [Evaluation Metrics](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.17xcoxilhdkq) [4](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.17xcoxilhdkq)1

5.5 [Decision Table](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.5x7ung3uz95s) [4](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.5x7ung3uz95s)2

5.5.1 [Extension Module](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.momb79vbmvv) [4](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.momb79vbmvv)2

5.5.2 Verification Module 43

5.6 [Integration Testing](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.rrwskhw2tuwl) [4](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.rrwskhw2tuwl)3

5.7 [User Acceptance Testing](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.gvky7hx0kdr) [4](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.gvky7hx0kdr)3

5.7.1 [Process of Acceptance Testing](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.pjvnma7n47fv) [4](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.pjvnma7n47fv)3

5.7.2 [Analysis of Gathered Information](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.9v2y8ulreijz) [4](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.9v2y8ulreijz)4

5.7.3 [Summary of Gathered Information](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.6hyy80lvwrxg) [49](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.6hyy80lvwrxg)

5.8 [Discussion](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.listl8fjs2h3) [4](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.listl8fjs2h3)5

5.9 [Conclusion](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.7euoe6c0cvo2) 46

[**CHAPTER 6**](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.p75colr7xfj2) **47**

[**CONCLUSION & SUGGESTIONS**](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.y308ctb7hc08) **47**

6.1 [Introduction](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.ocwkkt7nr9n8) 47

6.2 [Challenges](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.k0l5y5szspeo) 47

6.3 [Advantages of the Application](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.mkjq04ymma5i) 47

6.4 [Disadvantages of the Application](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.zamtjkftzc5z) 48

6.5 [Future Development Cross-Platform](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.tlnzoyg8k441) 48

6.6 [Conclusion](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.q68wxtu71mcj) 48

[**REFERENCES**](https://docs.google.com/document/d/12jrTZvtvavNBsFEhgOiti3DbBUPZY6OK2Zer4hNqJwM/edit#heading=h.bqagoditvtyl) **49**

**LIST OF FIGURES**

**Figure Page**

Figure 1.1 Fake News 1

Figure 2.1 Framework of literature review 6

Figure 2.2 Fake News Detector 12

Figure 2.3 Review 13

Figure 2.4 FakerFact: Fake News Detection 13

Figure 2.5 Review 14

Figure 2.6 Bullshit Detector 14

Figure 2.7 Review 14

Figure 2.8 Fake News Detection with Pinocchio Alerts 15

Figure 2.9 Review 15

Figure 2.10 Debunker - Detect Fake News 15

Figure 2.11 Review 16

Figure 2.12 No Bias News 16

Figure 2.13 Review 16

Figure 2.14 120 Lies Fact Check Amalgamator 16

Figure 2.15 Centr: Political Bias Detector 17

Figure 2.16 Review 18

Figure 2.17 NewsCracker 18

Figure 2.18 Review 19

Figure 2.19 Fake News Checker 19

Figure 2.20 Fact Check 19

Figure 3.1 Agile Method 22

Figure 3.2 Use case Diagram 24

Figure 3.3 Activity Diagram 28

Figure 3.4 Sequence Diagram 29

Figure 3.5 Class Diagram 29

Figure 4.1 Extension tab 31

Figure 4.2 Main Modal Window 32

Figure 4.3 Result Modal window (Real news) 33

Figure 4.4 Result Modal window (Fake news) 33

Figure 5.1 Dataset 37

Figure 5.2 Logistic Regression Model 38

Figure 5.3 MultinomiaLNB (naive\_bayes) Model 39

Figure 5.4 Decision Tree Classifier Model 39

Figure 5.5 PassiveAggressive Classifier Model 40

Figure 5.6 Random Forest Classifier Model 40

Figure 5.7 LinearSVC Model 41

Figure 5.8 Analysis of Gathered information 45

## **LIST OF TABLE**

**Table Page**

Table 2.1 Comparison table of similar system 20

Table 3.1 Written Use case of Select News 25

Table 3.2 Written Use case of Chrome Extension Enabled 25

Table 3.3 Written Use case of ML Module 26

Table 3.4 Written Use case of Cross Referencing 26

Table 3.5 Written Use case of Display Result 27

Table 5.1 Confusion matrix 42

Table 5.2 Extension Module Decision Table 42

Table 5.3 Verification Module Decision Table 43

Table 5.4 User Rating 45

## **LIST OF ABBREVIATIONS**

## **SDK -** Software Development Kit

**UML -** Unified Modeling Language

**ML -** Machine Language

**NLP -**  Natural Language Processing

**SVM -** Support Vector Machine

**CNN -** Convolutional Neural Network

**RNN -** Recurrent neural network

**KNN -** K-Nearest Neighbor

**VS Code -** Visual Studio Code

**SDLC -** Software Development Life Cycle

**LR -** Logistic Relapse

**PCFG** - Probabilistic context-free grammar

**TF-IDF**  - Term frequency-inverse document frequency

**PCFG -** Probabilistic context-free grammar

## **CHAPTER-1**

# **INTRODUCTION**

## 1.1 Introduction

In this era, the trend of fake news has escalated rapidly that benefits both economically and politically. It needs to be prevented, the first thing that needs to be done is to detect the fake news and address it. Any individual needs a good workable network and a device that supports social media platforms or websites and the user is good to go to the vulnerabilities of fake news.

A perpetrator can be anyone just using a social media network or a person that has a particular motive of political entities, economical entities, extremists, or a group of people with ulterior motives. Fake news can be different websites or portals dedicated to confusing the users with real websites, manipulating information, misleading it, and displaying it under the name of a real one. It doesn’t mean that everyone who is in a social media network is the victim of fake news but the most vulnerable people are those who have limited education and awareness to identify fake news.

This project was encouraged when the global pandemic of coronavirus (2020) began and the main topic all over the Internet was the origination and cures of the virus which led many perpetrators to manipulate the cures of disease and the fake news of vaccines.



Figure 1.1: Source: Fake News <https://www.bbc.com/news/52124740> (2021)

This project is intended to help to detect fake news by developing a machine learning program that identifies news sources that might have been producing fake news through a collection of written tasks and patterns of labeled real and fake articles. With the use of a corpus that labeled real or fake articles, a classifier is built to decide the information on the content. This model focuses on identifying from which source fake news originated based on multiple articles of that source and once a source is identified as a producer of fake news, all the upcoming articles that are published by this source are fake news.

## 1.2 Problem Statement

* Traditional textual analysis exceeds the limits of fake news detection emerging from different perspectives.

## The existing chrome extension still lags cross-referencing features.

* The core task of identifying fake news needs identifying the language(sentences or words) which is used to deceive the readers and classifying fake news by learning word-level meaning.

## 1.3 Objective

* Obj1: To explore the maximum number of ML algorithms to find the best algorithm for detecting fake news.
* Obj2: To develop fake news detector chrome extensions using the explore ML algorithm.
* Obj3: To correctly identify fake news and develop multiple cross-referencing to and truth checking through this chrome extension.

## 1.4 Scope

### 1.4.1 User Scope

* Users will be able to identify fake news while surfing the internet.
* Users will be able to see the rating of each website.
* Users will be able to cross-reference the selected article with trusted sites.

### 1.4.2 System Scope

* The system will be able to generate a rating for the selected site.
* The system will be able to detect the authenticity of the news
* The system will be able to list the trusted site containing the similar news article

### 1.4.3. Constraints

### Limitations of signal and highlight based techniques:

Different linguistics cues suggest that another prompt set should be intended for a prospective circumstance which makes it hard, sum up, cues and highlight designing strategies across various subjects and domains. Such methodologies consequently would require more human contribution in the design process, evaluation, and usage of these cues for detection.

#### Limitations of linguistic examination based techniques:

Although this type of technique is regularly considered to be better than cue-based strategies it shockingly still doesn't extract and completely abuse the rich semantic and syntactic information in the substance. E.g.: The N-gram approach is basic, notwithstanding, it can't demonstrate more confounded relevant conditions of the content. Syntactic highlights utilized alone are additionally less incredible than word-based n-grams and a shallow blend of the two would not be viable in catching the perplexing relationship.

#### Limitations of the deep learning-based techniques:

Fake news discovery is as yet a test even to profound learning techniques, for example, Convolutional Neural Network (CNN), Recurrent neural network (RNN), and so on, because the substance of phony news is arranged in a manner it looks like reality to deceive readers; and without cross-referring to and truth checking, it is regularly hard to decide veracity by text analysis alone.

#### Limits of existing feedback-based techniques:

The issue with existing feedback-based techniques (e.g.: Response User Analysis, Response text analysis, Temporal Pattern Analysis, Propagation Pattern Analysis, and Hand-designed analysis) is the sort of preparing information that models are being prepared on. It is typically a depiction of clients' reactions that are generally gathered after or towards the finish of the proliferation interaction when adequate reactions are free. This energizes and gives motivation to the diminished quality in an exhibition on early recognition utilizing prepared models when there are fewer reactions gathered. The techniques likewise cannot refresh their state-dependent on steadily accessible clients' reactions.

Limitations of existing intervention-based techniques:

Intervention-based techniques like (Decontamination, Network checking, crowd-sourcing, and User Behavior Modeling) will in general be harder to assess and try particularly in complex conditions where there are numerous related connections and transactions. Likewise, they may make restrictive suspicions about specific cases which limits their appropriateness. [(Sharma, Karishma; Feng, Qian; He, Jiang; Ruchansky, Natali, 2019](https://arxiv.org/abs/1901.06437))

## 1.5 Methodology

In this project agile model is used as a software development life cycle due to the following advantages:

* Individual and team interactions over processes and tools
* Working software over comprehensive documentation
* Customer collaboration over contract negotiation
* Responding to change over following a plan

1.6 Development tool

1.6.1 Machine Learning: This project utilizes machine learning models to confirm the article content and title. Utilizing machine learning, it had the option to check if the title is misleading content or not, choose if the article content is fake or real. In correlation with parsing, machine learning was utilized to take care of classiﬁcation issues. Then again, parsing was utilized for extracting data.

1.6.2 JavaScript:JavaScript is used to create chrome extensions since it is high-level, often just-in-time compiled, and multi-paradigm. It has curly-bracket syntax, dynamic typing, prototype-based object orientation, and first-class functions.

1.6.3 Python**:** This is the high-level python project of recognizing fake news and managing fake and genuine news. Utilizing sklearn, we fabricate a TfidfVectorizer on our dataset. At that point, we introduce a PassiveAggressive Classifier and fit the model. Eventually, the precision score and the disarray matrix reveal to us how well our model tolls.

1.6.4 VS Code: VS-Code is an advanced code editor and in this project, it has been used as a code editor to complete the project.

1.6.5 Draw.io: VS-Code is a high-level code editor and in this project, it has been utilized as a code editorial manager to finish the project. Draw.io is a free and online diagramming application that helps in building fundamental diagrams such as class diagrams, flowcharts, sequence diagrams, etc. of the entire project.

## 1.7 Conclusion

The problems of fake news and disinformation assume a significant part in these days’ life. This is a result of the advanced level of technology and specialized communication methods among people, we have empowered information spreading among individuals with no veriﬁcation.

This is a motivation behind why scientists began looking for answers to prevent fake news and disinformation from spreading without any problem. In any case, it is notable that controlling the ﬂow of information online is impossible.

# **CHAPTER - 2**

# **LITERATURE REVIEW**

## 

## 2.1 Introduction

In this chapter, there is a more detailed explanation about the background study of the existing system and analyzes the key features and the similar system with different systems. It also analyses the different algorithms used to build this system as well as the existing system’s algorithms.

There will also be a comparison report between different other chrome extensions. We also talk about SVM, KNN, NLP and ML, and other tools used for building this system. This chapter is all about literature preview and study about features of other systems. The table will contrast the proposed system and another similar system and clarify why the feature isn't accessible for the proposed system.

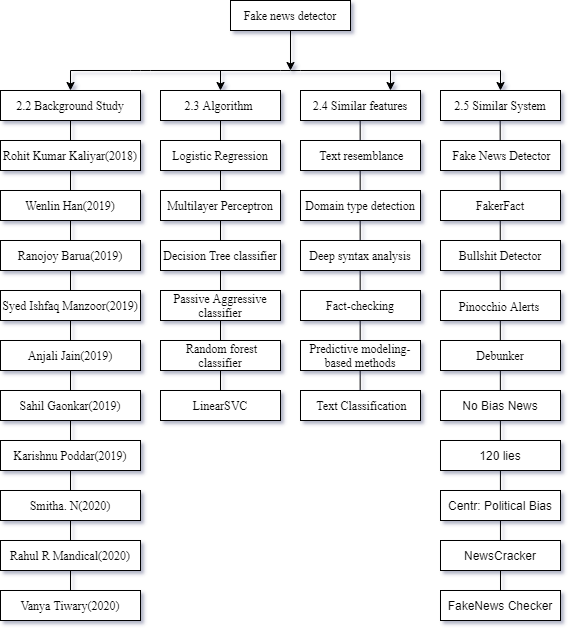


Figure 2.1: Framework of literature review

## 2.2 Background study

Identifying fake news by means of online media addresses a few new and testing research issues. Despite the fact that fake news itself isn't another issue nations or gatherings have been using the news media to execute exposure or effect errands for a significant long time the rising of web-created news by means of web-based media makes fake news an even more amazing force that challenges standard article principles.

The fakeness of news could emerge out of expanded points of view, which is past the limits of traditional textual analysis. For instance, fake news may result from its specific situations or speaker, and it may not straightforwardly identify with its contents. (Victoria L Rubin, Yimin Chen, and Niall J Conroy, 2015)

Fake News Detection Using a Deep Neural Network has utilized Natural language processing, Machine learning, and Deep Learning strategies to execute this model and look at which model will give more precise outcomes. Using the DGX1 Nvidia PC to get exact outcomes and divide the dataset into real and fake news. This model contains assortments of AI calculations, for example, K Nearest Neighbour (KNN), Naive Bayes (NB), Random Forest (RF), and Decision Tree (DT). Profound learning models, for example, Shallow Convolution Network (SCN) and Very Deep Convolutional Network (VDCN) and gated organizations, for example, Gated Recurrent Unit (GRU) with the assistance of Convolution Network and Long Short-term Memory (CN-LSTM). Moreover explored the sufficiency of word embeddings and word2vec features with Deep Neural Networks. This model uses chi2 for incorporation in the Machine learning model to deliver more precise outcomes. (Rohit Kumar Kaliyar, 2018).

Fake news discovery in social organizations utilizing machine learning and deep learning performance evaluation have alluded to some traditional machine learning approaches, for example, Deception modeling, clustering, Naive Bayes are assessed for accuracy detection TF-IDF and PCFG with Convolution neural organization and Recurrent neural network models are evaluated to consider the execution with ordinary machine learning methods. Deep Modeling talks in high space dimensional content and computer algorithms that should be removed to show sensibly. Fake data and fake articles have an extraordinary proportion of ordinary properties so to describe them Naive Bayes arrangement is applied. The precision of the framework is used utilizing different strategies, for example, bigram repeat which is used by TF-IDF and PCFG with the mix of the two techniques. The CNN and RNN models are alluded to for text mining or picture acknowledgment additionally RNN for time and game plan-based assumptions. (Wenlin Han, 2019)

An Application for Fake News Article Detection utilizing Machine Learning Techniques, created dependent on machine learning approaches, for example, long short-term (LSTM) and Gated recurrent unit (GRU) to portray data into spam or original. The preliminary outcomes on the dataset arranged in this work look precise. The model is moreover arranged and taken at other data and comparable results show the efficiency of the proposed model. To distinguish the working of the FNAD model. It is attempted and used on 1,000 content data in FNAD's information to evaluate the execution of the model using assessment estimations, for instance, the Confusion model, Score model, and Accuracy model. Execution aftereffects of the proposed model utilizing FNAD produce 80.2% precision. Subsequently, it very well may be used to affirm text data from various sources of the web preceding enduring it as a reality. (Ranojoy Barua, 2019)

Their sorts of Fake News portrayed depend on Visual, User, Knowledge, Style, Stanze, various strategies for Fake News Detection, for example, Linguistic-Basis, Clustering, Predictive, Content-CueBased, Non-Text-Cue-Based modeling. Algorithms considered for the proposition were Naive Bayes, Decision Trees, SVM, Neural Networks, Random Forest, XG-Boost. Numerous techniques for the more precise results of detection of Fake News, for example, Convolutional Neural Networks, Deep Boltzmann Machine, Deep Neural Network, and Deep Autoencoder Model were applied on this which were convenient and yielded more precision in perceiving whether the News is mis-driving or Real. (Syed Ishfaq Manzoor, 2019)

A Brilliant System for Detection of Fake News of which announced about Stopping of Fake News by Facebook and Whatsapp as well, Aggregators, News Authenticators, News Suggestion or Recommendation System help in advancing the Fake News dependent on User interest. The proposed system was checked and assessed by various usage strategies like NLP, Naive Bayes, CNN, "Guileless Bayes SVM NLP" and found the precision as 76%, 74%, 87%, 94% separately and discussed every single execution and its assessment results and System Architectures, SVMs' and so on. Also, as of high precision, to utilize SVM, NLP with Naive Bayes Algorithm usage which yielded the most noteworthy accuracy on their emergency. (Anjali Jain, 2019)

Detection of Online Fake News A Survey. This research incorporates a model that groups conflicting data into veritable or counterfeit news by enrolling a score dependent on various data obtained from URLs. This survey uses Machine learning techniques like Linear Regression, Logistic Regression, and Support Vector Machine (SVM) with the assistance of Multilayer perceptron (MP) to acquire more exact information. All information that is gained is then added to get the last score and classify the data as fake or real. From this overview, it is seen that the given data is preprocessed utilizing preprocessing systems like tokenizer and stemming. TF IDF utilizes Context-free grammar (CFG) for additional order of information. From breaking down the overview we can express that recognizing fake data on social sites is simpler than some other online platforms. (Sahil Gaonkar, 2019)

A Comparison of Various Machine Learning Models for Accurate Detection of Fake News. The main point of this model is to deal with the issue of fake data classification. With the scores of unmistakable vectorizers to be explicitly checked and term frequency-inverse document frequency (TF-IDF) is differentiated to find the vectorizers. English stop words are used for improving the score and precision. Various classifications are utilized: Regression, Support Vector Machine (SVM), and Decision tree classifiers are used for exact fake source recognizable proof. The reenactment data show that the Support Vector Machine and TF-IDF will give the most exact expectation. The popular technique is to distinguish word tally utilizing the TF-IDF strategy and afterward, the estimations of TF-IDF are joined to each word archive. The Text preprocessor changes numbers over to messages previously and is used with TF-IDF for the identification of fake content. (Karishnu Poddar, 2019)

ML classifiers for Fake news recognition and its accuracy. Count-Vectorizer, TF-IDF Vectorizer, Word Embedding for the estimation of best accuracy and best execution. By using classification algorithms the precision obtained with SVM straight characterization calculation with TF-IDF Vectorizer highlight extraction is 94% exact. The utilization of Neural networks has comparable accurate outcomes than SVM direct characterization calculations. Utilizing Neural organizations with direct grouping calculations would make the order increasingly intricate, hence the technique for SVM over Neural Networks is less unpredictable. The first stage in the proposed framework is text assortment alluded to as datasets. The subsequent advance is text pre-handling in which transformation steps are incorporated. The third is Feature extraction which incorporates the encoding of words. Fourth is Classifiers of News which incorporates SVM, Logistic Regression, Decision Trees, and so forth Order with TF-IDF Vectorizer yield more exact outcomes than Count-Vectorizer and Word Implanting measure, Therefore TF-IDF Vectorizer for Classification. (Smitha. N, 2020)

The execution of Detection of Fake news utilizing Machine Learning by utilization of Naive Bayes and Passive-Aggressive classifiers. These classifiers are utilized to take care of models dependent on TF-IDF Vectorizer where TF-IDF represents Term Frequency-Inverse Document Frequency of which worth increments with an increase in the number of times the word appears in the report, however, there is a misfortune in the semantic significance of words. The Passive-Aggressive Classifier is a Linear-based model, Tokenizers for contributing to Deep Neural Network Models (DNN). Contemplated Keras vectorizes and changes over content corpus into vectors or arrangement of numbers which is passed to the macintosh pool layer to ascertain a solitary greatest incentive for every one of the info channels. After the assessment of classifiers, it was seen that Naive Bayes and Passive-Aggressive Classifiers with DNN have beaten a large portion of the Datasets. In this way, decide on the DNN with these referenced classifiers. (Rahul R Mandical, 2020)

A grouping of News features utilizing Machine Learning, Clickbait, Propaganda, Comment or Opinion, and Satire or Humor are the arranged sorts of Fake News, to Count Vectorizer, TF-IDF Vectorizer, Hashing Vectorizer. The usage of Logical Regression, Decision Tree, Random Forest with various Vectorizers gave the outcome that utilization of Logistic Regression calculation with TF-IDF Vectorizer has more exact outcomes when contrasted with different calculations with TF-IDF Vectorizer. More than 73% of feature Fake News was precisely identified and ordered, while Decision Tree was exact up to 62% and Random Forest up to 66% separately. The examination done by them was effective in characterizing the HeadLines of News. By this proposition, Sorted out that a Logical Regression calculation with TF-IDF Vectorizer would yield the best outcomes for the characterization of Headline in the News. (Vanya Tiwary, 2020)

### 2.3 Algorithm

This project will use the following algorithm to create an ML module and to evaluate the performance of fake news detection classifiers.

2.3.1 Logistic Regression

As we are characterizing text dependent on a wide rundown of highlights, with a binary output (true/false or true article/fake article), a logistic relapse (LR) model is used, since it gives the characteristic condition to bunch issues into equal or various classes. We performed hyperparameter tuning to get the best result for all individual datasets, while different limits are attempted before acquiring the most extraordinary exactnesses from the LR model.

Logistic regression utilizes a sigmoid capacity to change the output to a probability value; the goal is to limit the expense capacity to accomplish an ideal likelihood.

2.3.2 Support Vector Machine:

Support vector machine (SVM) is another model for binary classification issues and is open in various bits capacities. The objective of an SVM model is to assess a hyperplane (or decision limit) in light of a rundown of capacities to assemble data centers. The segment of the hyperplane changes according to the number of features.

As there could be various freedoms for a hyperplane to exist in an N-dimensional space, the task is to perceive the plane that separates the data motivations behind two classes with the most maximum margin.

2.3.3 Multilayer Perceptron

A multilayer perceptron (MLP) is an artificial neural network, with an info layer, in any event, one covered layer, and a yield layer. MLP can be essentially just about as clear as having all of the three layers; regardless, in our assessments, we have changed the model with various limits and different layers to make an ideal envisioning model.

2.3.4 Passive-Aggressive classifiers

The passive-aggressive algorithms are a family of algorithms for large-scope learning. They are like the Perceptron in that they don't need a learning rate. Be that as it may, in spite of the Perceptron, they incorporate a regularization boundary. In this project, a Passive-Aggressive Classifier will be used.

2.3.5 Decision Tree classifier

A decision tree is a tree-like design whereby an inward node addresses an attribute, a branch addresses a decision rule, and the leaf nodes address a result. This works by dividing the data into separate segments as indicated by an attribute selection measure, which for this situation is the Gini index (in spite of the fact that we can change this to information gain if we needed). This basically implies that we each split points to reduce Gini impurity which estimates how impure a node is as to incorrectly classified outcomes.

2.3.6 Random Forest Classifier

Random Forest is an ensemble method that joins different decision trees to characterize, so the result of random forest is normally better compared to decision trees. Random forests is a supervised learning algorithm. It very well may be utilized both for classification and regression. It is additionally the most flexible and simple to utilize algorithm. A forest is composed of trees. It is said that the more trees there are, the stronger a forest is. Random Forests makes decision trees on randomly selected data samples, gets forecast from each tree and chooses the best solution through voting. It additionally gives a decent indicator of the feature significance.

## 2.4 Common feature of the similar system:

2.4.1 Text resemblance: Other chrome extensions compare different articles and compare the news titles and find out the text resemblance using NLTK.

2.4.2 Domain type detection: Other chrome extensions use chatNoir API and search news titles of different articles in web snapshots.[(Ermakova, Liana & Seffih, Hosni & Firsov, Anton & Le Noé-Bienvenu, Guillaume & Guibon, Gaël., 2019)](https://halshs.archives-ouvertes.fr/halshs-02391141)

2.4.3 Deep syntax analysis: The deep syntax analysis can be inspected using Probabilistic context-free grammar (PCFG) and these are used by different systems to perceive fake news. Accentuation structures are portrayed by changing sentences into parse trees. Things, activity words, etc are changed into their syntactic constituent parts. Probabilities are permitted to the parse tree. This system perceives the standard arrangements like lexicalization and parent center points, etc it perceives misleading with 85-91% precision, dependent upon the order used in the assessment.

2.4.4 Fact-checking: Fact-checking is a kind of "data-based examination of fake news" that revolves around assessing the trustworthiness of data used by different systems. There are two sorts of fact-checking, to be explicit: manual and automatic.

### 2.4.5 Predictive modeling-based methods: The detection of fake news is moreover cultivated through prescient demonstrating-based procedures used by different systems. One sort would be the determined backslide model. In this model, positive coefficients increase the probability of truth while negative ones increase the probability of misleading.

2.4.6 Text Classification: Text classification otherwise called text tagging or text categorization is the way toward arranging text into coordinated groups. By utilizing Natural Language Processing (NLP), text classifiers can consequently investigate text and afterward assign a group of pre-characterized labels or classifications dependent on its content. It is used in fake news detection as it does sentiment analysis, topic detection, and language detection, and analyses thousands of texts in just a few seconds.

### 

## 2.5 Research of similar systems

2.5.1 Fake News Detector

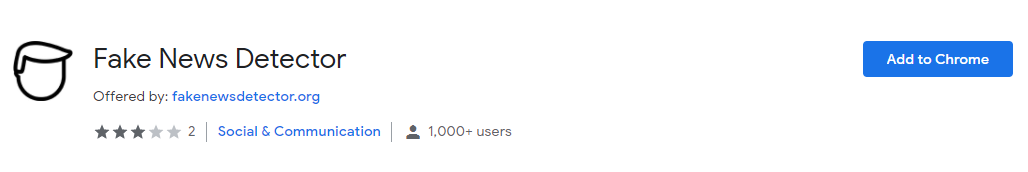
****

Figure 2.2: Fake News Detector (2021)

**1000+ users**

Detect Fake News, Click Bait, and more from news sources, and flag it to help other people.

The Fake News Detector permits detection and flag news straightforwardly from Facebook and Twitter into Legitimate, Fake News, Click Bait, Extremely Biased, Satire, or Not news.

Subsequent to flagging another story, others that have the expansion will actually want to see your flagging, will focus harder on it, and may signal. The information is then secured on a data set and scanned by their robot, Robinho.

Robinho peruses the data given by users and learns with time to naturally flag news as Fake News, Click Bait, and so forth, in light of its content. And then, even new news that nobody saw might be immediately flagged. The augmentation at that point shows on Facebook the assessment from others and the robot.

****

Figure 2.3: Source: Review (Fake News Detector, n.d.)(2021)

### 2.5.2 FakerFact: Fake News Detection

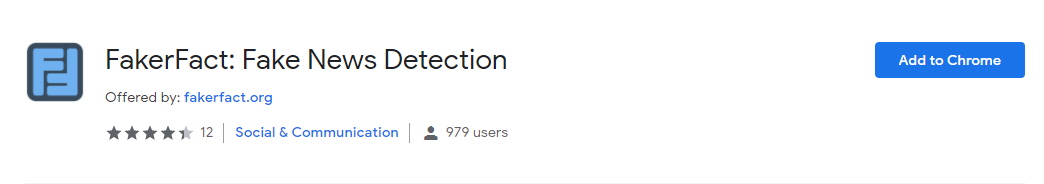
****

Figure 2.4: FakerFact: Fake News Detection (2021)

**979 users**

This tool will open up another tab on www.fakerfact.org to check the current tab's substance for indications of fake news. With such a great deal of information coming at you reliably, it might be hard to figure out what real Journalism is Fake News. While PCs can't make reference to you what's real, etc, they can help give information that will help you make that confirmation.

FakerFact uses a machine-learning calculation called Walt (named after Walter Cronkite). Walt has sought after a considerable number of articles from districts wherever on the web and has been set up to recognize material Fake News plans.

For example, Walt can tell whether the site page we are seeing offers characteristics of articles that are average of uplifting news projecting, assessment pieces, misdirecting content, intrigue thoughts, or spoof.

By then this detector makes an affirmation of whether it believes that the article is an authentic and solid news source or if it is Fake News.

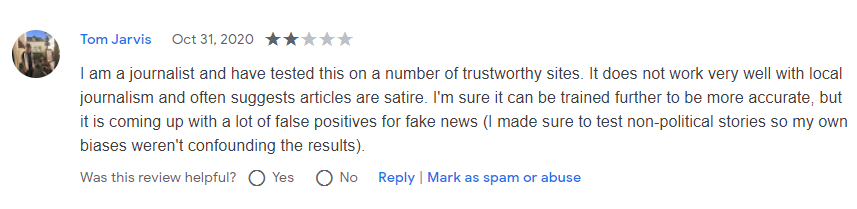


Figure 2.5: Source: Review (FakerFact: Fake News Detection, n.d.)(2021)

### 2.5.3 Bullshit Detector

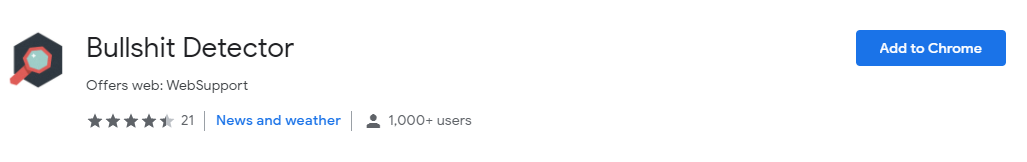


Figure 2.6: Bullshit Detector (2021)

At the point when we visit a site with untrusted content, it distinguishes the page as conceivably unsafe. Bullshit Detector is associated with the information base of the conspirators.SK project. The quarrelsomeness of the substance is surveyed by a specialist commission.

The extension means to caution against false reports, conspiracy ideas, scams and hatreds, and fascist ideologies.

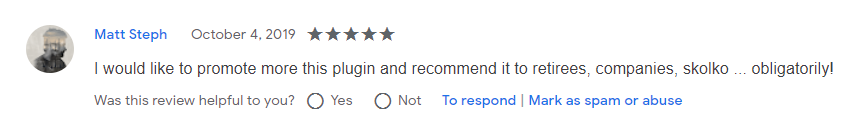


Figure 2.7: Source: Review (Bullshit Detector, n.d.)

2.5.4 Fake News Detector with Pinocchio Alerts

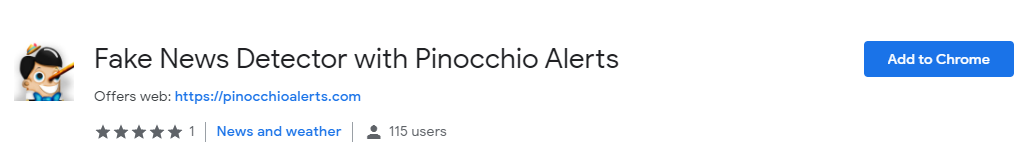


Figure 2.8: Fake News Detector with Pinocchio Alerts

Fact check module to identify fake news in articles we read. On the off chance that it has been accounted for it will tell us. Pinocchio Alerts is an expansion that cautions us if the site we are perusing has been accounted for as off base (due to fake news identified with that site).

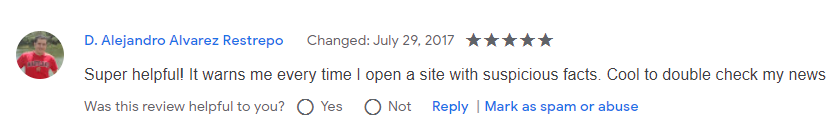


Figure 2.9: Source: review (Fake News Detector with Pinocchio Alerts, n.d.)

2.5.5 Debunker - Detect Fake News

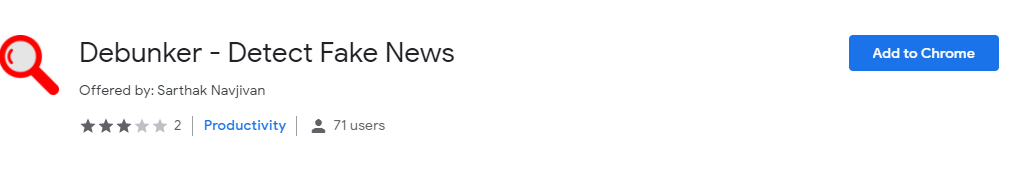


Figure 2.10: Debunker - Detect Fake News

Detects fake news sites utilizing advanced web scraping technology and a machine learning algorithm.

A debunker is a digital tool that can accurately survey the quality of any online news source.

In contrast to current arrangements, which essentially analyze the title of the site to decide if the source is reliable or not, Debunker utilizes Artificial Intelligence to break down the real content of the site to check whether it is trustworthy.

On the snap of its icon, Debunker utilizes progressed web scratching advances to accumulate data from the site the client is on and breaks down a few of its key variables, like the article's writer, distributor, area name, composing intricacy, picture sources, prominence, utilization of catchphrases, and that's only the tip of the iceberg.

At long last, utilizing a neural organization AI model, Debunker loads and registers these variables to decide if the source is tenable.

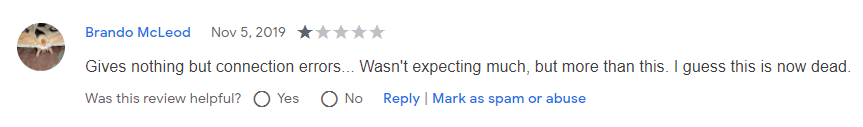


Figure 2.11: Source: Review (Debunker - Detect Fake News, n.d.)

2.5.6 No Bias News

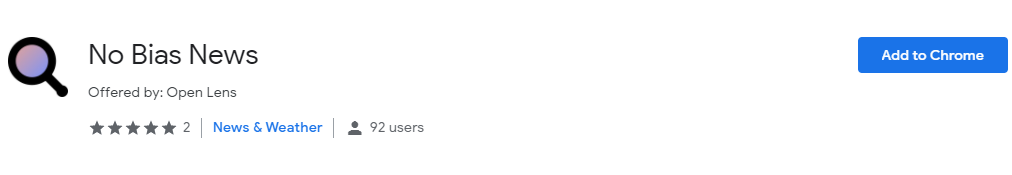


Figure 2.12: No Bias News

No Bias News uncovers the biases and backgrounds of different news sources.

The No Bias News Chrome expansion takes the inclination and "factual reporting" evaluations from Media Bias Fact Check and shows these in a coordinated way, alongside other applicable data, when visiting the particular news site.

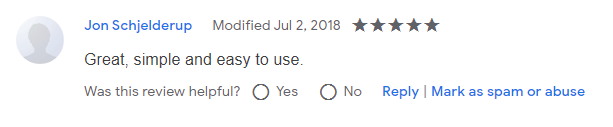


Figure 2.13: Source: Review (No Bias News, n.d.)

2.5.7 120 Lies: Fact-Check Amalgamator

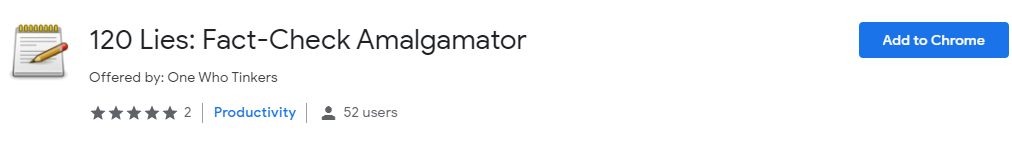


Figure 2.14: Lies: Fact-Check Amalgamator

Flags 120 already fact-checked false or misleading proclamations

Explanations were found from Snopes, PolitiFact, FactCheck.org, and that's only the tip of the iceberg. Generally political, for the most part, USA-related, some metropolitan legends.

Accepting misrepresentations is a difficulty that traverses ideological groups. All assertions with sources: goo. gl/zmoY7V.

2.5.8 Centr: Political Bias Detector



Figure 2.15: Centr: Political Bias Detector

Centr is a Chrome Extension that permits you to comprehend the bias of articles you read. This bias isn't attached to the site, yet to the real content of the article!

Centr checks the article you are perusing and produces a Political Bias score from an assortment of measurable and profound learning strategies. The diagram has 5 classes along the x-pivot: Left: Strong liberal bias Left-Leaning: Slightly liberal bias Center: Congrats! The news source you are reading is fairly objective & unbiased. Keep reading! Right Leaning: Slight conservative bias Right: Strong conservative bias Scores address the article just, not simply the general article of the news source itself.

The Centr model was prepared on more than 110,000 articles from more than 15 distinctive media sources across the political range. This information was enhanced with extra articles gathered through web scratching. It utilizes the intelligence of the groups - we source prominent sentiments for locales that are one-sided in a specific way, similar to All Sides and Media Bias Fact Check. Utilizing AI and ML to add between the articles, it teases out predisposition from the content of the text itself!

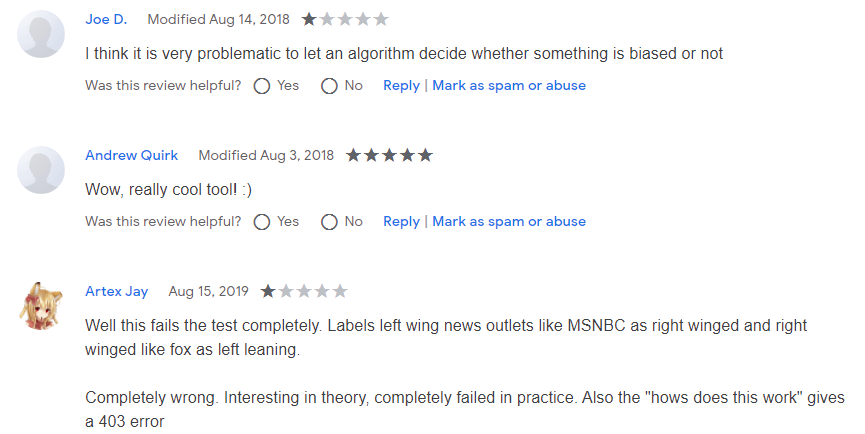


Figure 2.16: Source: Review: (Centr: Political Bias Detector, n.d.)

2.5.9 NewsCracker

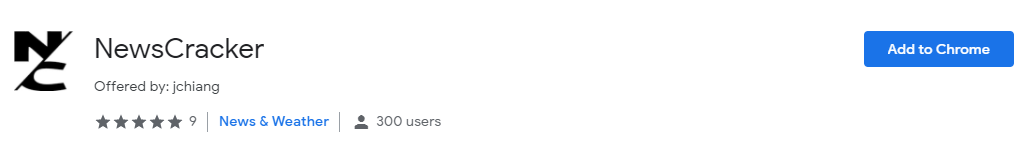


Figure 2.17: NewsCracker

NewsCracker by checking an article for expected mistakes or predispositions, you can avoid fake news and stay an insightful, educated news shopper while perusing a web progressively overwhelmed by deceiving content.

On the off chance that you can't help contradicting a rating on an article you're seeing, you can "banner" the rating, which will incite us to investigate the rating and see what turned out badly.

You may likewise tell it of any bugs or mistakes that surface, either through the expansion or by messaging it at newscrackerco mpany@gmail.com, so we can keep on improving the client experience.

Kindly note that its scores are produced by a calculation, and keeping in mind that it gives a valiant effort to ensure that they appropriately mirror the substance of each article, they are not total pointers of exactness or error, similarly as they are not outright markers of subjectivity or objectivity.



Figure 2.18: Source: Review: (NewsCracker, n.d.)

2.5.10 FakeNews Checker



Figure 2.19: FakeNews Checker

This extension gives context menu things that permit a client to check news bias of connections and pictures. For instance on Facebook, you can without much of a stretch figure out what kind of left or right-inclining inclination a site has, as logged by mediabiascheck.com.

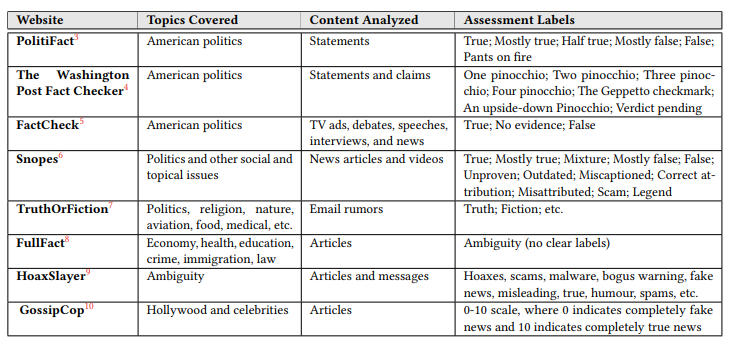
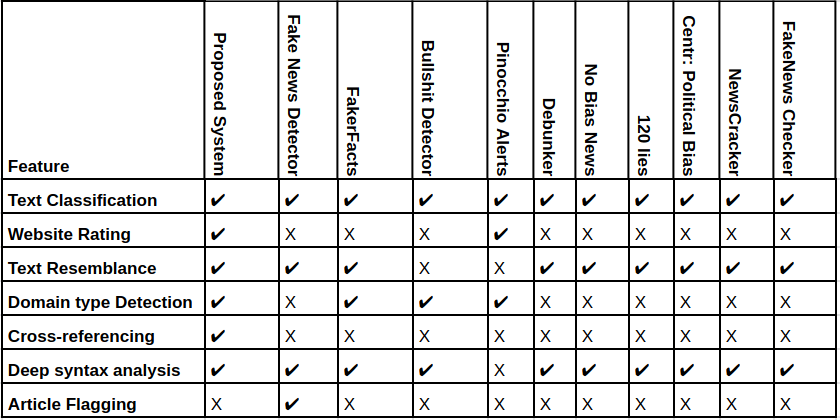


Figure 2.20: Fact Check Source: (Zhou & Zafarani, 2018)

2.6 Comparison Table

Table 2.1: Comparison Table



2.7 Conclusion

In this chapter, we compared different fake news detector chrome extensions and studied their features detailly. Also learned about the background study of fake news detectors today and the algorithms being used by different chrome extensions.

Researched similar features of other chrome extensions and the reviews from other users as well. Also did a comparison study of different fake news detectors. This system is worked with complete NLP (Natural Language Processing) and comparison based on text resemblance and domain type. Conclusion the Existing chrome extension does not have cross-referencing features.

To deploy this system Python is utilized as a programming language, NLP and ML are utilized to prepare and construct the models, CNN is utilized for picture manipulation and to distinguish fake news reports.

# **CHAPTER - 3**

# **METHODOLOGY**

## 

## 3.1 Introduction

In this chapter, there is a detailed explanation of the methodology that has been utilized to assemble this system. First and foremost, the agile model is utilized as a software development life cycle (SDLC) to build up this system. Other than that there are various sorts of unified modeling language (UML) diagram configuration to propose the system.

The diagram that has been utilized in this project is a Use case diagram where it will clarify the usefulness that can be executed by the actor and the agent. Moreover, there will be an activity diagram to show the behavioral diagram to picture the work process of the sequence activity and the activity expected to play out the undertaking. At that point, there will also be the sequence diagram which is additionally an interaction diagram to show the relationship between the object and the time sequence.

Finally, there will be a class diagram that portrays the design of the system by understanding the system class and the class attribute, activity, and relation among the object to give clear modeling that can translate the programming code.

## 3.2 Agile Model

Agile is a methodology by which a team can execute a project by splitting it into many steps and requiring continual communication with partners at any point and continuous development and iteration. The Agile approach continues with consumers explaining how to use the final product and what challenges it can fix.

The reason to use agile methodology on this application are:

1. By being adaptable, quick, lean, productive, and reliable, Agile Methodologies have supplanted the traditional waterfall model methodologies.
2. The Agile strategy follows best practices that contribute rapidly to excellent software.
3. Agile methodologies require an incessant audit in a trained manner, which subsequently builds the qualities of leadership to augment coordination.
4. In a competitive world, agile methodologies are tried and end up being exceptionally strong by reacting to the adjustments in the business.

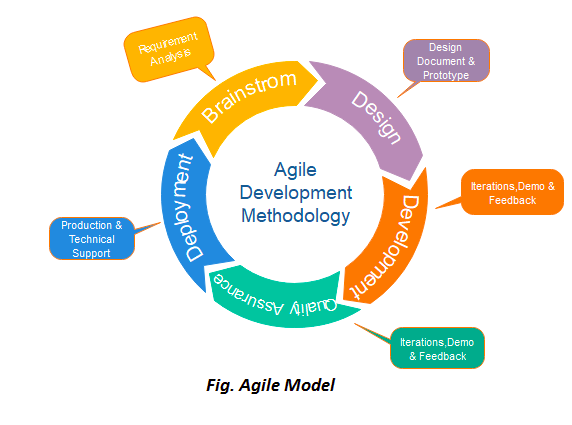


Figure 3.1: Source:https://www.javatpoint.com/software-engineering-agile-model (2021)

The project is divided into different phases; every phase has its own property.

### Phase 1

Brainstorm (Requirement analysis)

Before the execution of the planning of an application, it needs to make the underlying documentation that represents the initial prerequisite. Some of them are:

* The final objective of the project will be reached.
* Distinguish the issue with the current system.
* Distinguish the system advantage and issues that can happen.
* Analyze a comparable system.

Phase 2

Design (Design document and prototype)

In the design of a system, there are two ways to deal with architecture: one is a visual depiction i.e. graphic design and the other is the framework's architectural framework. Our system design will be simple and the user will interact less with it.

Software Design

During the principal emphasis of software design, the prerequisites were made and attempted to handle the requirement and propose the tools expected to accomplish the best outcome.

UI/UX Design

During this SDC stage, the UI model plan of the site is made. On the off chance that the system satisfies the client's desires, the clients will review and the UI and client experience meets the plan head and brilliant principle then the plan will be last for the following stages.

Phase 3

Development (Iteration, Demo, and criticism)

During this stage, the development of an application is begun. The application is prepared to compose code and convert the design documentation into a real fully functional application. This phase is generally the longest as it is the foundation of the entire cycle.

Phase 4

Quality Assurance (Iteration, Demo, and criticism)

This phase is to test the system and make the system without bugs and viable with the gadgets. During this stage, the system is tested dependent on the system's principal objective.

Phase 5

Deployment (Production and technical support) in this phase, the last chrome augmentation is delivered and appropriated to the client to chip away at the new highlights and tackle the bug looked at by the client. Finally, the audit of the chrome augmentation is gathered from the client.

## 3.3 Use case Diagram

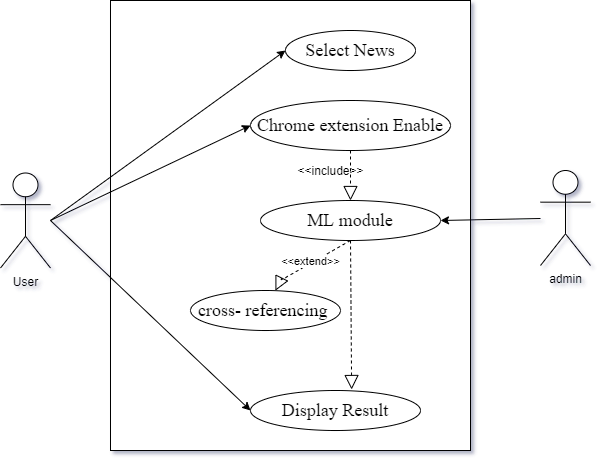
****

Figure 3.2: Use Case Diagram

## 3.4 Written Use Case

## 3.4.1 Select News

Table 3.1: Written Use Case of Select News

|  |  |
| --- | --- |
| Use Case ID: | UC-1 |
| Use Case Name | Select News |
| Actors: | Users |
| Precondition | Chrome Extension enabled |
| Postcondition | Fetch Article from users |
| Trigger | Users need to select the article |

|  |
| --- |
| Main Success Scenario |
| -Users provide news or article  -Users need to install a chrome extension |
| Extensions |

3.4.2 Chrome Extension Enabled

Table 3.2: Written Use Case of Chrome Extension Enabled

|  |  |
| --- | --- |
| Use Case ID: | UC-2 |
| Use Case Name | Chrome extension enables |
| Actors: | Users |
| Precondition | Article insert in extension |
| Postcondition | Fake or Real article |
| Trigger | Detect button pressed |

|  |
| --- |
| Main Success Scenario |
| -System logo is shown in the extension tab |
| Extensions |

3.4.3 ML Module

Table 3.3: Written Use Case of ML Module

|  |  |
| --- | --- |
| Use Case ID: | UC-3 |
| Use Case Name | ML Module |
| Actors: | Users |
| Precondition | Article Fetched |
| Postcondition | Fake or Real article |
| Trigger | Detect button pressed |

|  |
| --- |
| Main Success Scenario |
| -Module detects whether the input article is real or fake |
| Extensions |

3.4.4 Cross-Referencing

Table 3.4: Written Use Case of Cross Referencing

|  |  |
| --- | --- |
| Use Case ID: | UC-4 |
| Use Case Name | Cross-referencing |
| Actors: | Users |
| Precondition | Authentication of the article |
| Postcondition | Similar article display |
| Trigger | Detect button pressed |

|  |
| --- |
| Main Success Scenario |
| -The article is real, the System should provide the same/similar article from a trusted site list. |
| Extensions |

3.4.5 Display Result

Table 3.5: Written Use Case of Display Result

|  |  |
| --- | --- |
| Use Case ID: | UC-5 |
| Use Case Name | Display Result |
| Actors: | Users |
| Precondition | Article input from the user |
| Postcondition | Result displayed |
| Trigger | Detect button pressed |

|  |
| --- |
| Main Success Scenario |
| * The system displays the result i.e. This article is fake or real * If the article is real, the System should provide the same article from a trusted site. |
| Extensions |

## 

## 3.5 Activity Diagram

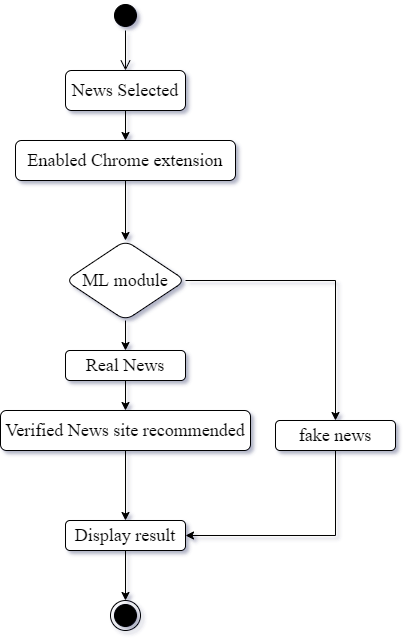
****

Figure 3.3: Activity Diagram

## 3.6 Sequence Diagram

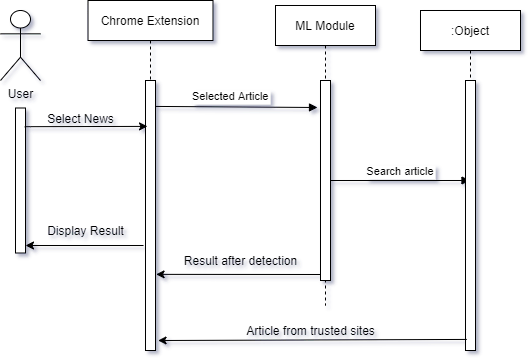


Figure 3.4: Sequence Diagram

## 3.7 Class Diagram

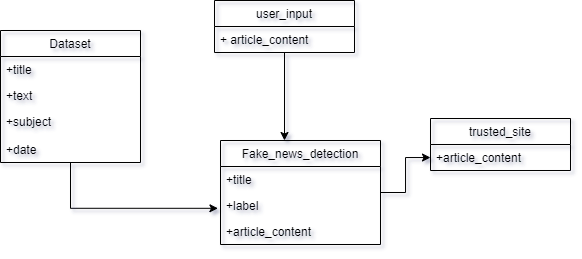
****

Figure 3.5: Class Diagram

## 3.8 Conclusion

Overall the fake news detector is developed using agile methodology. Agile is used in this model due to its flexibility, clear documentation and development speed, and quality due to sprint. The UML design properly displayed the required steps to implement this system, Activity diagram showed the behavioral diagram to picture the work process of sequence activity and the activity expected to play out the undertaking. As well, a sequence diagram is there which has shown an interaction diagram to show the relationship between the object and the time sequence. The class diagram portrays the design of the system by understanding the system class and the class attribute, activity, and relation among the object to give clear modeling that can translate the programming code. This project will make use of a real-world dataset. The datasets are obtained from Kaggle. Thus, the development method of this project is represented by the diagrams.

**CHAPTER 4**

**INTERFACE**

## 4.1 Introduction

This chapter explains all the activities an end-user can perform in the system with a brief explanation of each functionality. It efficiently explains the whole functionality in a step-by-step process. The whole system is presented through the end-user view, who wants to check the authenticity of a certain news article. The results are displayed through the alert window. Since the system is a chrome extension, the UI is minimal and focused on enhancing the browsing experience without distracting from it.

## 4.2 Interface explain

### 4.2.1 Extension Tab

Rather than drawing the user's full attention to the extension, the entire system is designed to optimize the user's productivity. The Extensions tab is located at the top-right of the browser window (between the Address Bar and the Menu Button).

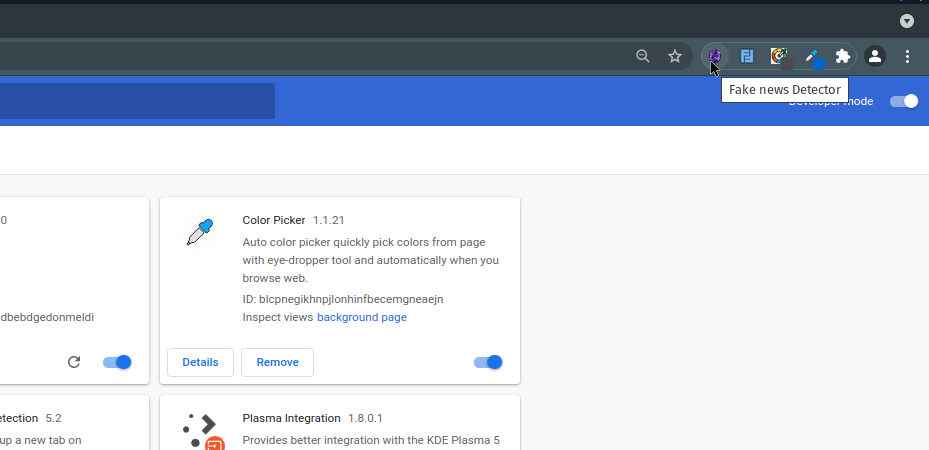


Figure 4.1: Extension Tab

As shown in the above Figure 4.1. The extensions are on the top-right of the browser window. Users can click on the icon. The extension modal window popup appears when the user clicks on the little icons that indicate the extension which is shown in the below figure 4.2

### 4.2.2 Main Modal Window

The main modal window is a modal window that is triggered when the users click on the icons on the extension tab that indicate the fake news extension.

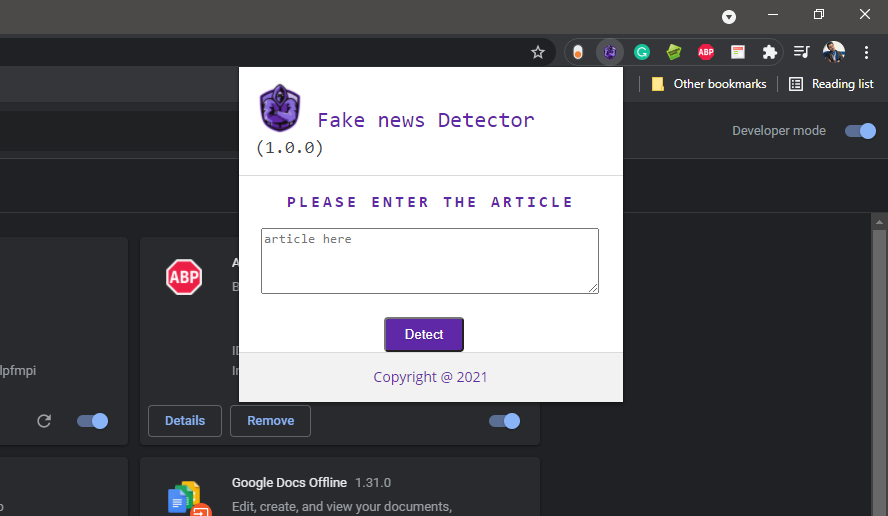


Figure 4.2: Main Modal Window

In this figure 4.2, users can see that the system requires the URL of the article the users want to verify. After providing the article URL when users click on the “Detect” button the article URL is further processed to bring out the desired result, to determine whether the article is fake or real. Below the “Detect” button, there is the footer where copyright is shown.

### 4.2.3 Result Modal Window (Real News)

Result modal window is a modal window that shows the outcome of whether the selected news article is fake or real.

In figure 4.3, Users can see the result of the article that the user entered through the Main modal window. As the result displays “This article is real” in the most subtle way. Below the result being displayed are lists of trusted sites with similar articles which are notified to the user. Users can copy the provided URL or click on the visit button which redirects them towards the page of the URL or to the particular website.

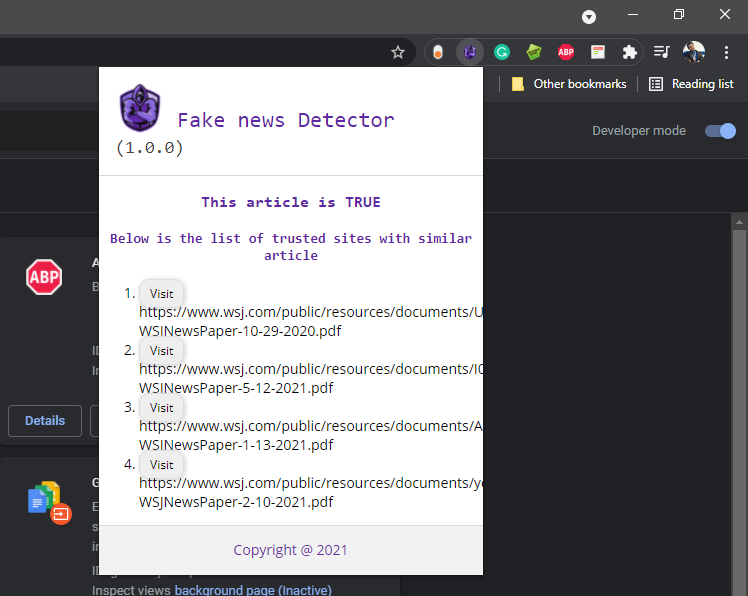


Figure 4.3: Result Modal Window (Real News)

### 4.2.4 Result Modal Window (Fake News)

Result modal window is a modal window that displays the outcome determining whether the chosen news article by the user is fake or real.

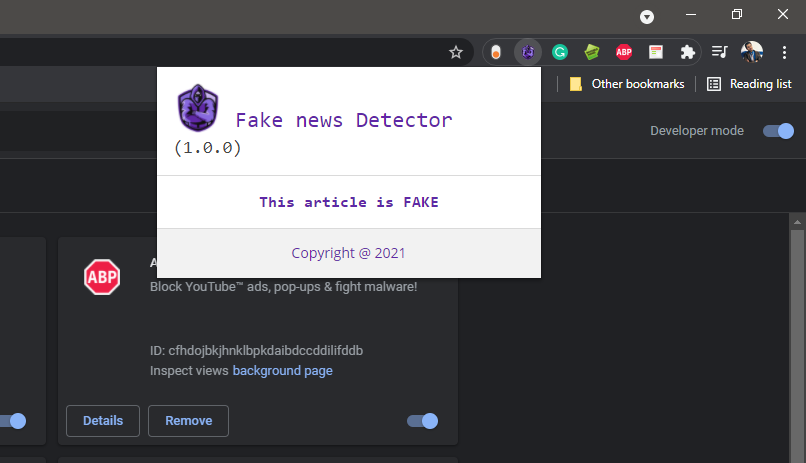


Figure 4.4: Result Modal Window (Fake News)

In the given figure 4.4, Users are shown the outcome of the article that they entered through the Main Modal window. This shows the result "This article is Fake". As the article of a specific site is set apart as fake, then that site is evaluated through a rating model and kept in a database for future references.

## 4.3 Conclusion

The overall interface has satisfied the given requirement and objective stated in chapter One. Some of the actions that are to be done by users are displayed through a modal window. Every extension’s interface is developed based on the common design concept with the consistency of color, title tab design, validation testing, reviews as well as promotion to reduce the short-term memory load by navigation of tab icons. The UI is designed to optimize the user's workflow rather than to draw all attention of the user to the extension.

# **CHAPTER 5**

# **TESTING AND RESULT**

## 5.1 Introduction

Software testing is an important aspect of software development where it evaluates the quality of software products and services provided. Software testing helps to prevent bugs, improve performance, reduce development costs and prevent unsatisfied clients. It allows an understanding of the risk of software implementation.

Software Testing is fundamental since, supposing that there are any bugs or glitches in the software, these can be identified early and can be corrected before the software product is conveyed. Appropriately checked software products keep up with reliability, stability, and good execution, which likewise brings about time investment funds, cost-productivity furthermore, client dependability.

* Unit Testing
  + Testing in which the whole software is divided into different unites and segments then these units or segments are tested individual. The goal is to ensure that the output of each unit of the product code is as expected.
  + Unit help to reduce costs by revising bugs early in the development cycle.
  + The phase of the decision table for this research is highlighted below:
    - To discover and work on the significant condition needed for the project.
    - For API integration of this extension, Postman is used for testing.
    - Fill out the Boolean condition of the conditions and activities for each test situation. Each experiment should reflect all potential conditions performed by the users and actions for applications. At last, survey the judgment table to make sure the responses to each experiment are right.
* Integration Testing
  + Testing is directed to distinguish flaws in the system and the interface.
  + This testing is performed when the API is incorporated into the extension application to test the outcome of the API result.
  + This testing is utilized when the web API is utilized to contact the ML model.
  + This testing is performed when the application is delivered to the end-user.
* User Acceptance Testing
  + User Acceptance Testing (UAT) is a sort of test directed by the end-user or the developer to verify/accept the software system before moving the software program to the production environment.
  + This testing is finished by giving the delivered beta variant of the application to gather the client feedback, client experience and to handle the client expectation just as API response time when utilizing the application by different clients simultaneously.
* Static Analysis
  + Static research examines requirements, objectives, spectrum, and design without running software.
  + This analysis is performed depending on the review and the end-user acceptance testing.

## 5.2 Dataset

The selected ML algorithms were trained by using a Kaggle dataset. Which has 1,061 Total Votes. Containing 23,481 fake news and 21,417 real news articles. The size of the Data where the models were trained is 110 MB. The dataset contained 4 columns i.e. title, text, subject, and date of collection. Since there were two datasets real.csv and fake.csv it was easy to label them as true and fake. Only the title column was used for the main features (criteria). The dataset was divided into two parts i.e. training data and testing data, 75 % of the data were used for training the model and 25% for the testing.

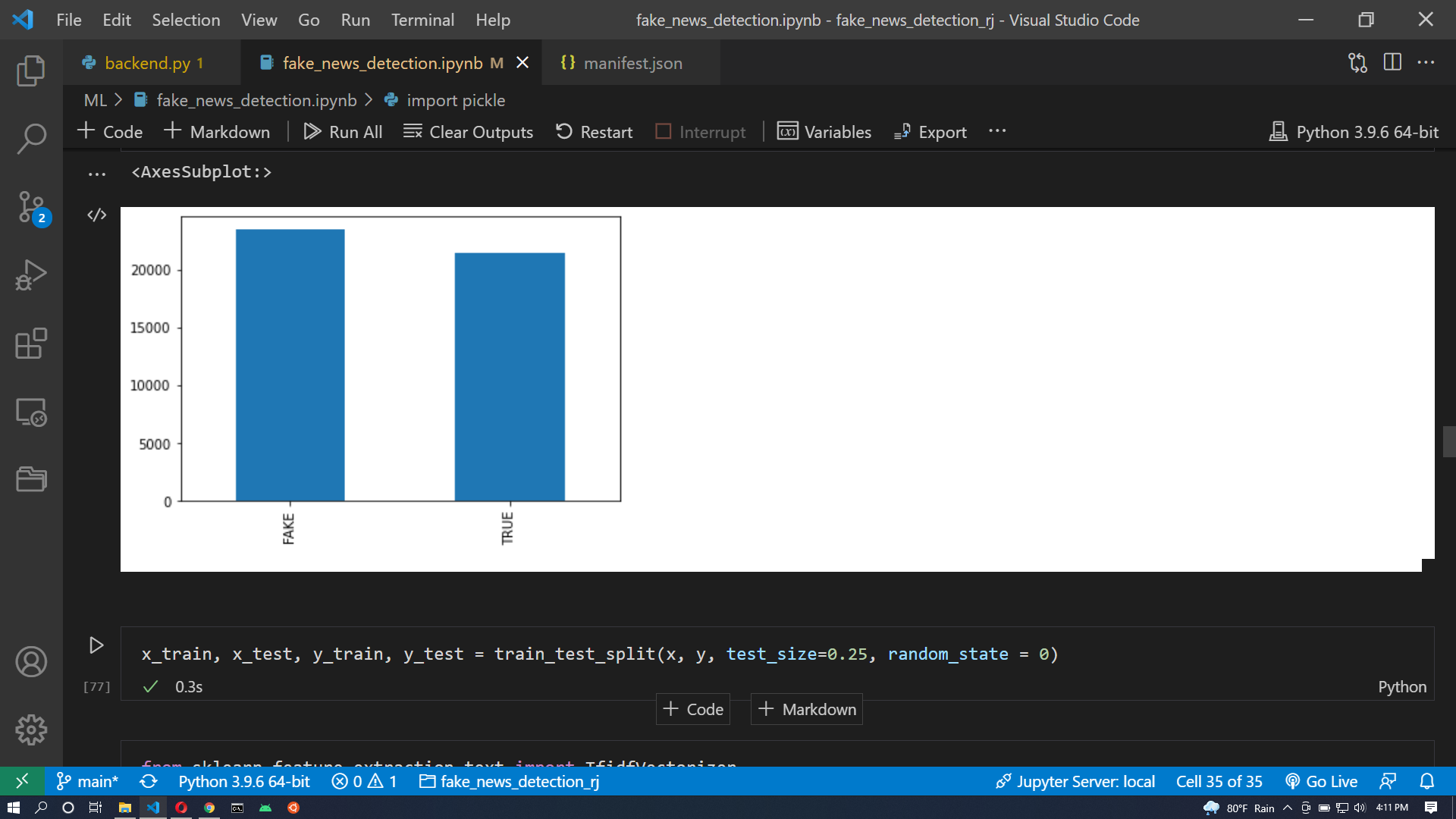
**

Figure 5.1: Dataset

As shown in the above Figure 5.1, the total data is divided into two labels i.e. fake and true. Where it contains 23,481 fake news articles and 21,417 real news articles.

## 5.3 ML Model Results

The model was trained using Vscode JupyterNotebook. The dataset collected from Kaggle is pre-processed to remove all the stopwords, Stopwords are the words in the language which does not add much meaning to the sentence for this regular expression library. After that data is vectorized, in this process of vectorization the words are converted in vector form as Vector algebra can be used to detect patterns/relationships in data. To make the corpora of documents more appealing to computers, it must be translated into a numerical format. Bag of Words was one of the techniques used to accomplish this. With the help of TF-IDF, we represent the corpus in matrix form, where the weight given to each token is determined not just by its frequency in a document, but also by how frequently that phrase appears across the corpora. After vectorization, we divided the data into test and train sets for the ML models. In this project, the data is used to train and test six different ML models.

### 5.3.1 Logistic Regression Model

The accuracy of the logistic regression model was 94.9042316258352, Approx~95%

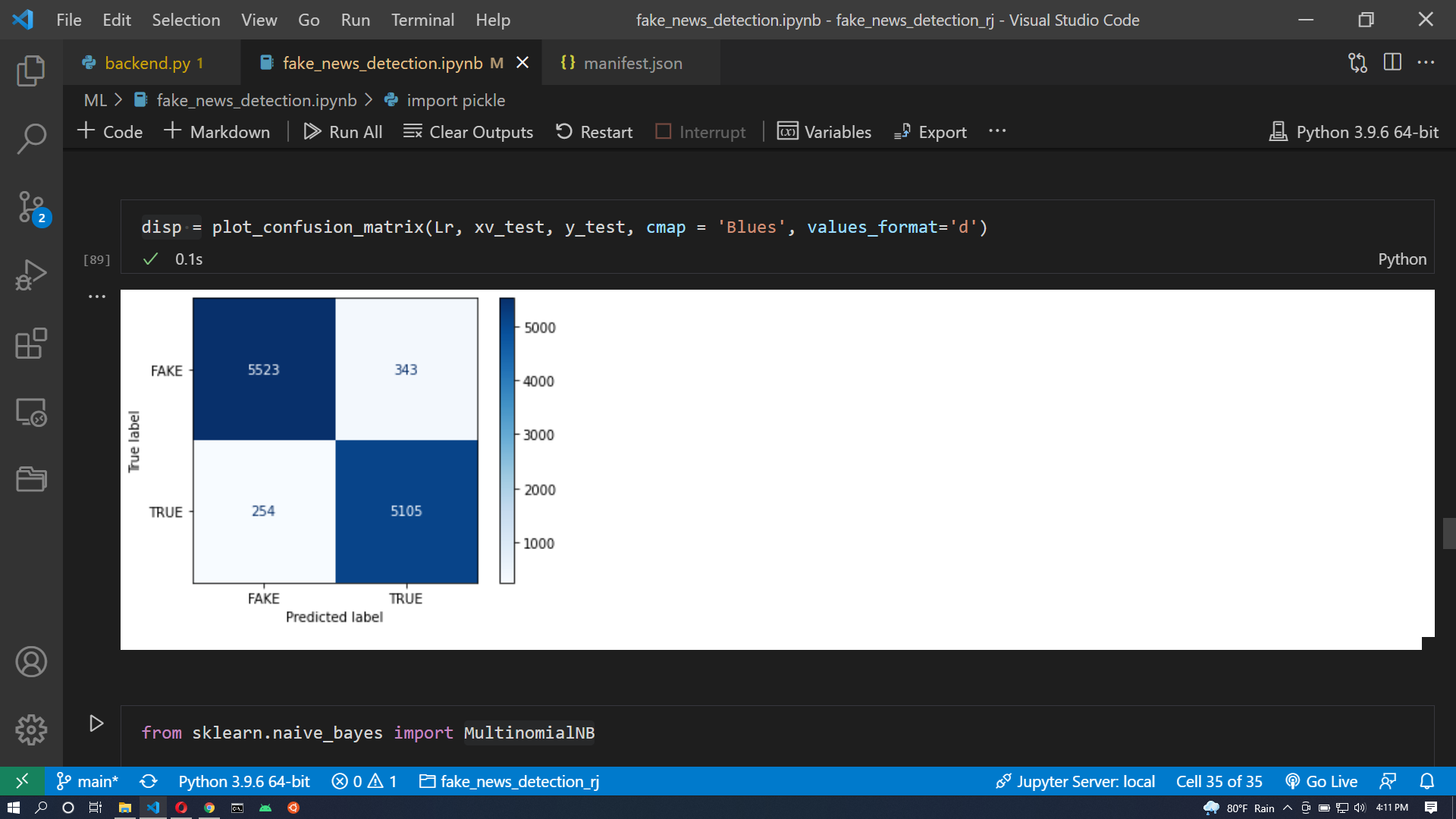


Figure 5.2: Logistic Regression Model

The above figure: 5.2, is the result of the confusion matrix of Logistic Regression, A confusion matrix is a table that describes a classification model (or "classifier's") performance on a set of test data with known actual values. The true labels are plotted in the y-axis and the predictions made by the model are plotted in the X-axis. The true positives (TP) is 5105, true negatives (TN) is 5523 whereas, the false positives (FP) is 343 and false negatives (FN) is 254. True positives mean the value which the model predicted is true and is real. True negatives are the values that the model predicted are fake and are fake. Whereas, the false positives (FP) are the values the model predicted are real but are fake. It is also known as Type I error. And the false negatives (FN) are the values the model predicted as fake but are real news. This error is also known as Type II error.

### 5.3.2 MultinomiaLNB (naive\_bayes) Model

The accuracy of the naive\_bayes model was 93.9064587973274, Approx ~ 94%.

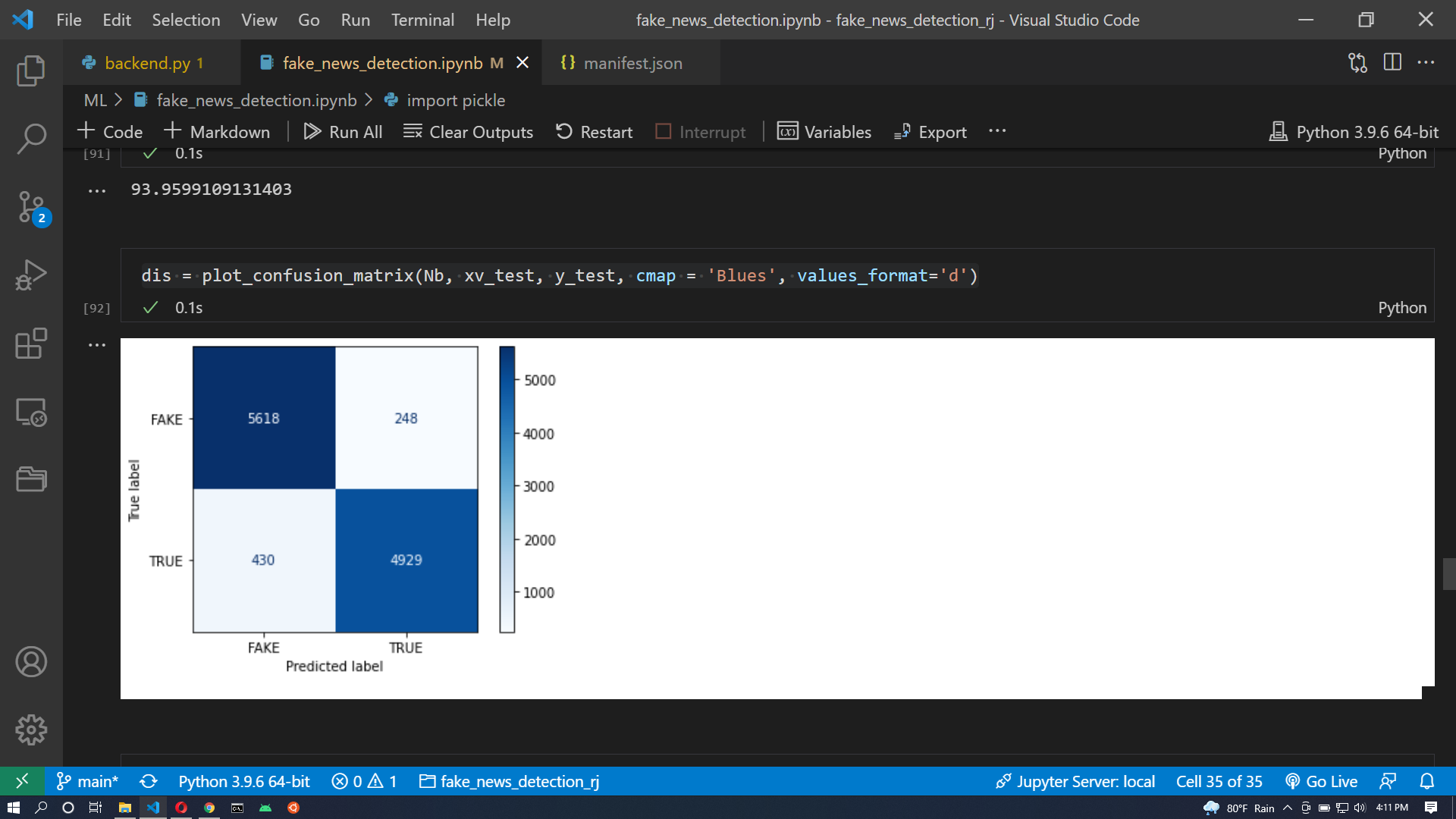


Figure 5.3: MultinomiaLNB (naive\_bayes) Model

The above figure: 5.3, is the result of the confusion matrix of the Multinomial (naive\_bayes) model where the true positives (TP) is 4929, true negatives (TN) is 5618 whereas, the false positives (FP) is 248 and false negatives (FN) is 430.

### 5.3.3 Decision Tree Classifier Model

The accuracy of the Decision Tree classifier model was 89.89755011135857, Approx ~ 90%

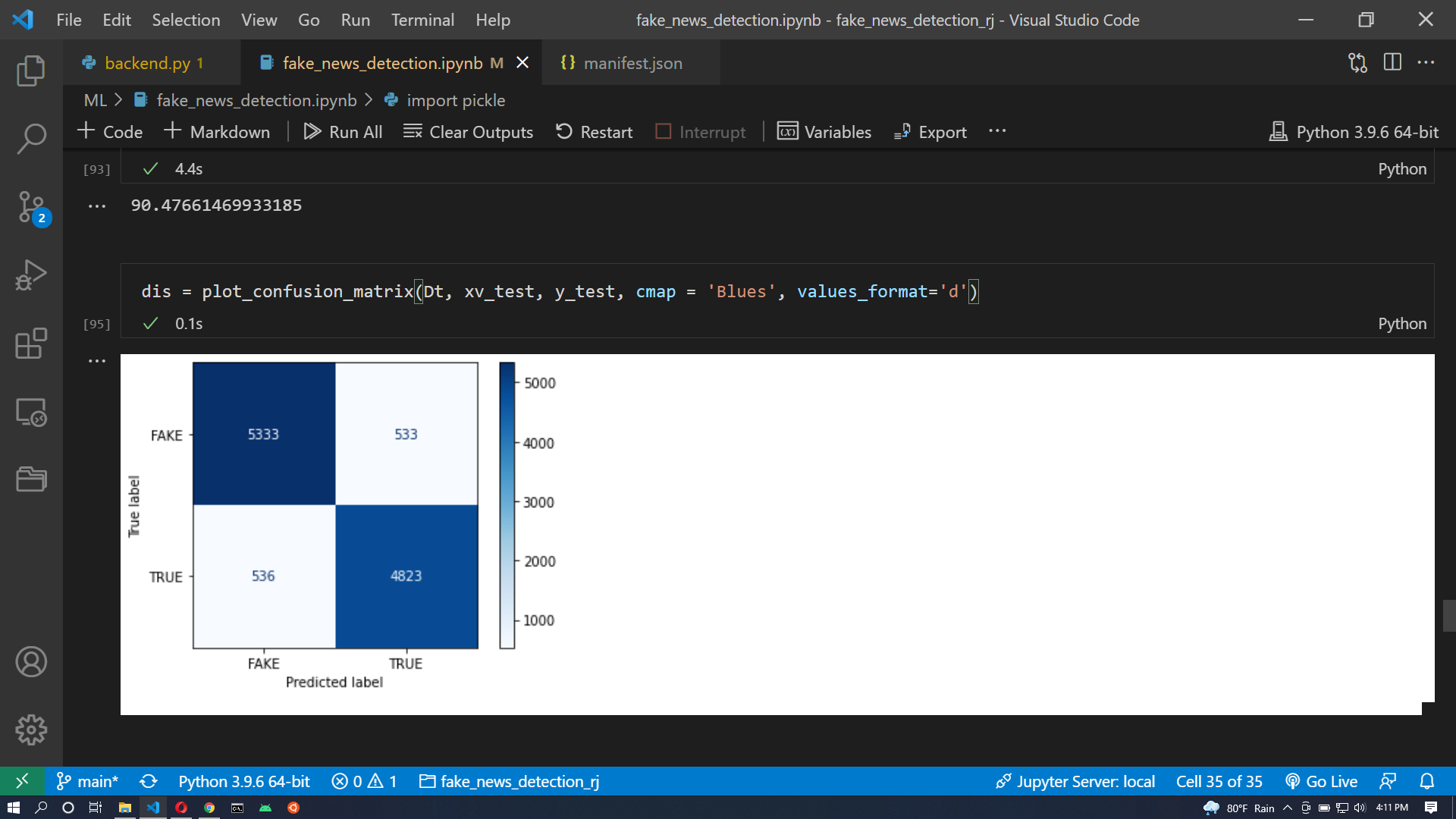


Figure 5.4: Decision Tree Classifier Model

The above figure: 5.4, is the result of the confusion matrix of the Decision Tree classifier model where the true positives (TP) is 4823, the true negatives (TN) is 5333 whereas, the false positives (FP) is 533 and false negatives (FN) is 536.

### 5.3.4 PassiveAggressive Classifier Model

The accuracy of the PassiveAggressive classifier model was 94.85077951002228, Approx ~ 95%

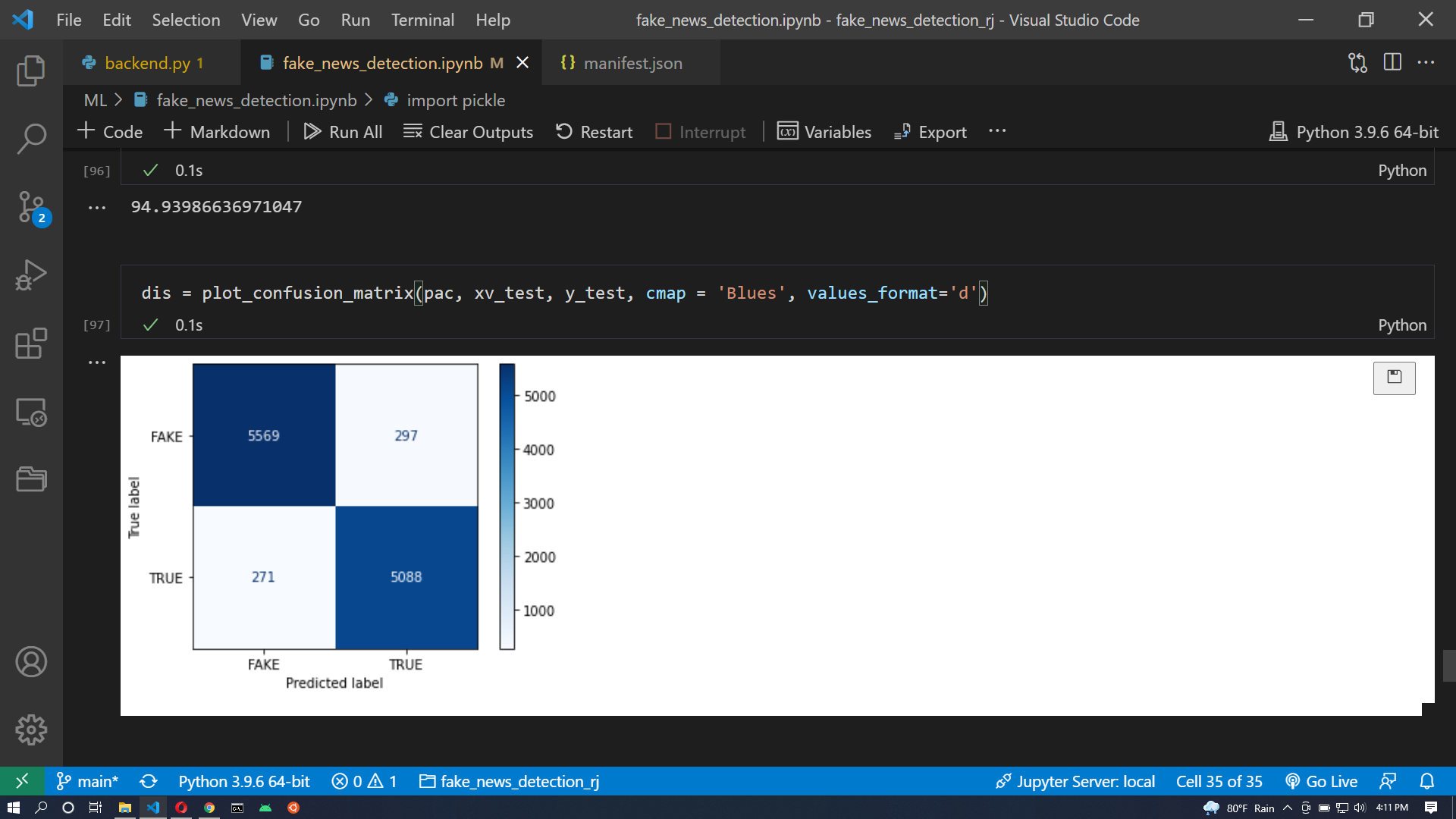


Figure 5.5: PassiveAggressive Classifier Model

The above figure: 5.5, is the result of the confusion matrix of the PassiveAggressive classifier model where the true positives (TP) is 5088, the true negatives (TN) is 5569 whereas, the false positives (FP) is 297 and false negatives (FN) is 271.

### 5.3.5 Random Forest Classifier Model

The accuracy of the Random forest classifier model was 94.56570155902004, Approx ~ 95%.

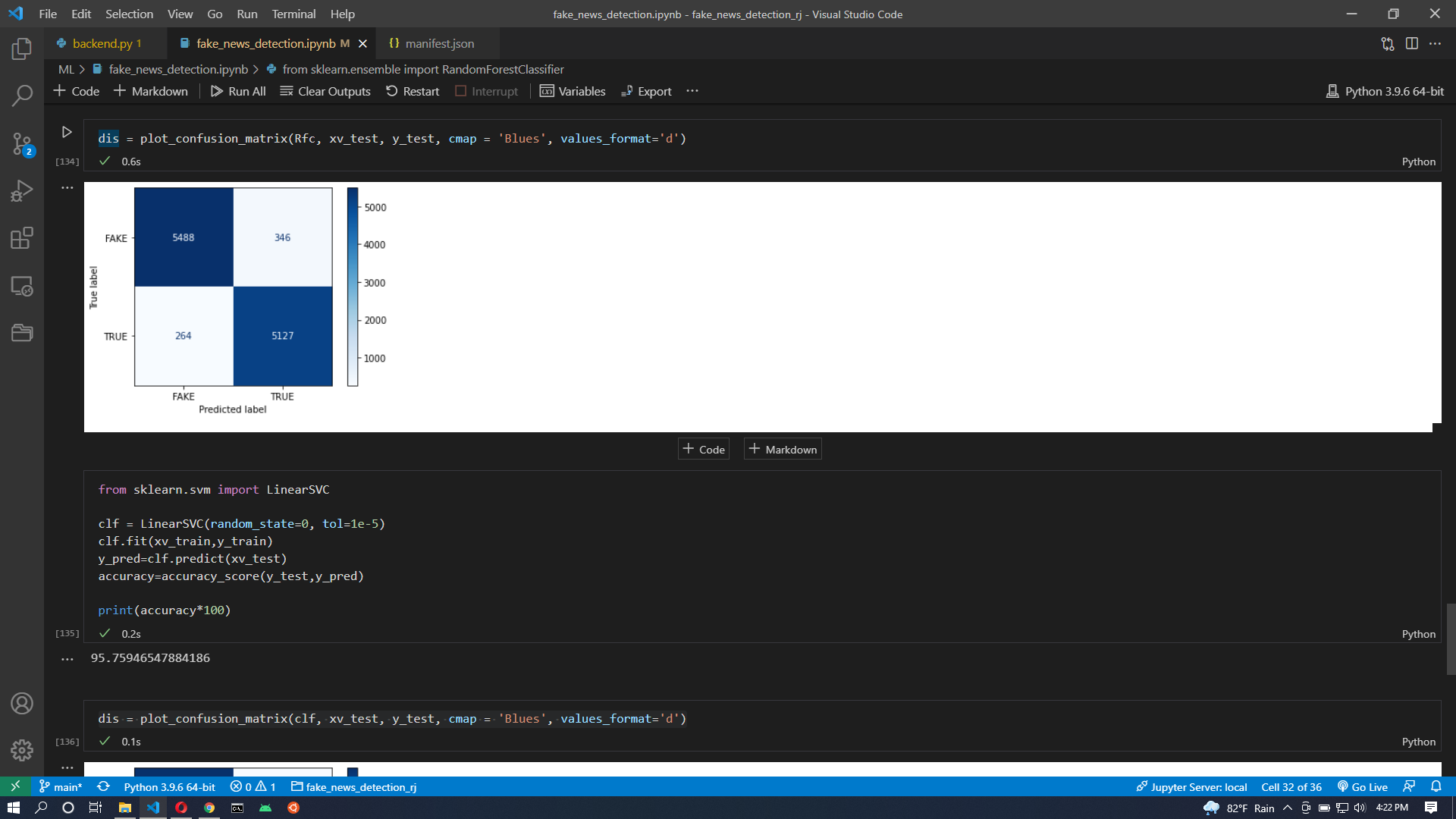


Figure 5.6: Random Forest Classifier Model

The above figure: 5.6, is the result of the confusion matrix of the Random forest classifier model where the true positives (TP) is 5127, the true negatives (TN) is 5488 whereas, the false positives (FP) is 346 and false negatives (FN) is 264.

### 5.3.6 LinearSVC Model

The accuracy of the LinearSVC model was 95.75946547884186, Approx~ 96%.

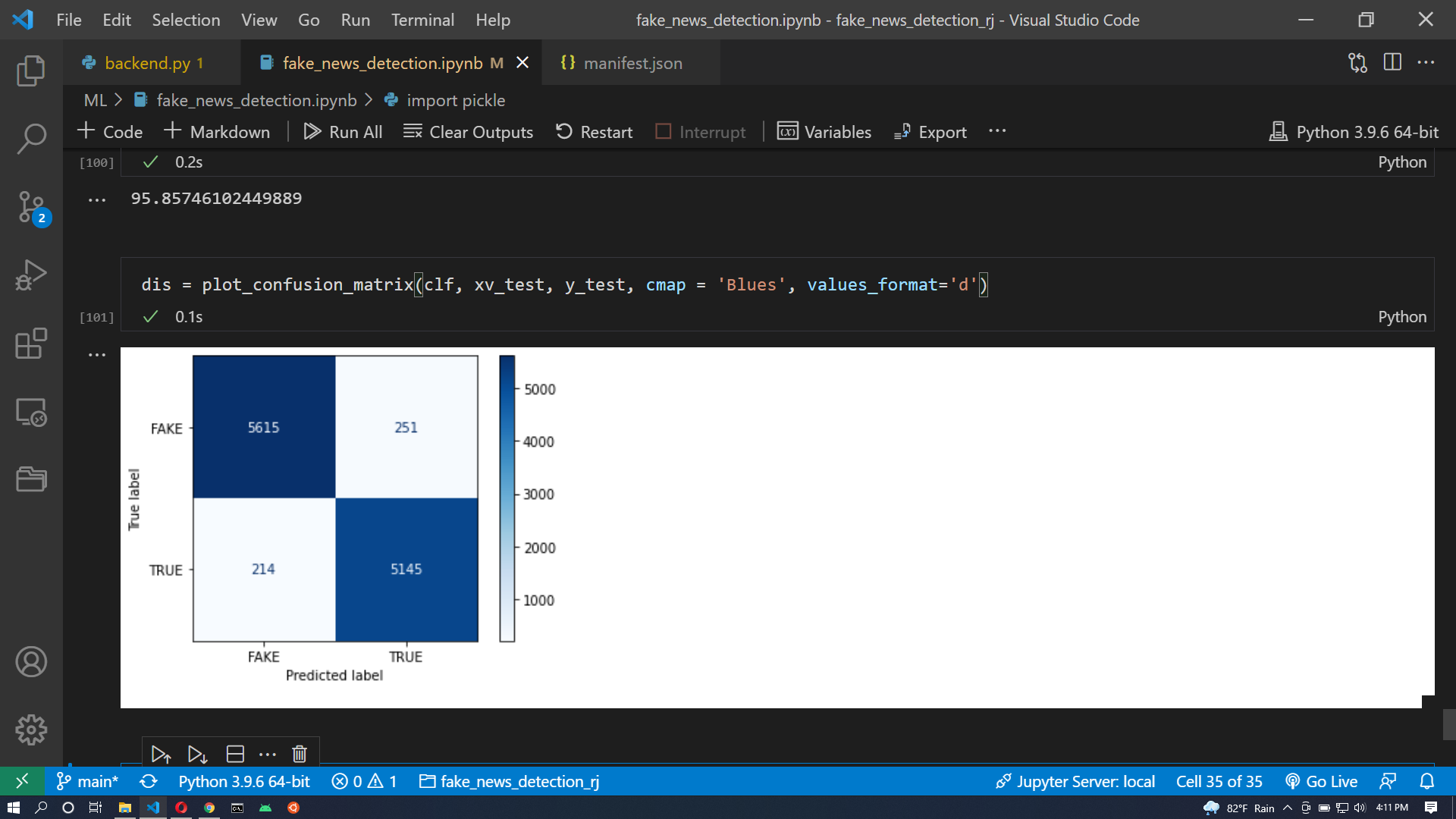


Figure 5.7: LinearSVC Model

The above figure: 5.7, is the result of the confusion matrix of the LinearSVC model where the true positives (TP) is 5145, the true negatives (TN) is 5615 whereas, the false positives (FP) is 251 and false negatives (FN) is 214.

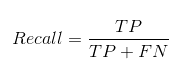
### 5.4 Evaluation Metrics

The output of the different models in the confusion matrix is highlighted below:

The ratio of correct predictions to total predictions made is accuracy. The formula is highlighted below:

Equation_Accuracy

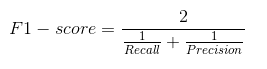
Recall indicates how many of the actual positive cases our model was able to properly anticipate.



Precision indicates how many of the instances that were correctly predicted turned out to be positive.

Confusion Matrix Precision

In practice, increasing the precision of our model reduces recall and vice versa. The F1-score combines both trends into a single number:



Because the F1-score is a harmonic mean of Precision and Recall, it provides a composite picture of these two metrics. When Precision equals Recall, it reaches its peak.

Table 5.1: Confusion matrix

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Accuracy | recall | precision | f1-score |
| Logistic regression model | 0.95 | 0.95 | 0.94 | 0.95 |
| Naive\_bayes Model | 0.94 | 0.92 | 0.96 | 0.94 |
| Decision Tree Classifier model | 0.91 | 0.91 | 0.90 | 0.91 |
| PassiveAggressive Classifier Model | 0.96 | 0.96 | 0.96 | 0.96 |
| Random Forest Classifier Model | 0.95 | 0.95 | 0.94 | 0.95 |
| LinearSVC Model | 0.96 | 0.96 | 0.96 | 0.96 |

## 5.5 Decision Table

### 5.5.1 Extension Module

Table 5.2: Extension Module Decision Table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Condition | Cases | | | |
|  | 1 | 2 | 3 | 4 |
| Click on the extension tab | T | T | T | F |
| End-user Input Response | T | F | T | F |
| Detect Button | T | T | F | F |
| Action | Cases | | | |
| 1 | 2 | 3 | 4 |
| Display result | T | F | F | F |
| Display verified site list | T | F | F | F |

### 5.5.2 Verification Module

Table 5.3: Verification Module Decision Table

|  |  |  |
| --- | --- | --- |
| Condition | Cases | |
|  | 1 | 2 |
| Result = “TRUE” | T | F |
| Action | Cases | |
| 1 | 2 |
| Display verified site list | T | F |

## 5.6 Integration Testing

For the Integration testing, the chrome extension frontend part sends a GET request with the article to check, to the API on the backend which is created in localhost using a python programming language. The article is processed and the predicted value with verified sites is fetched back to the frontend part from the API.

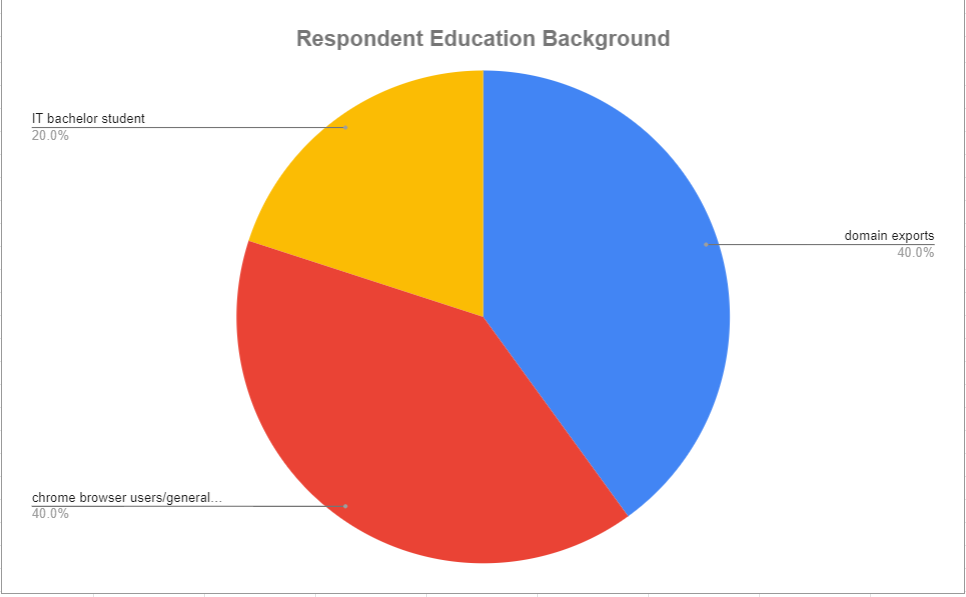
## 5.7 User Acceptance Testing

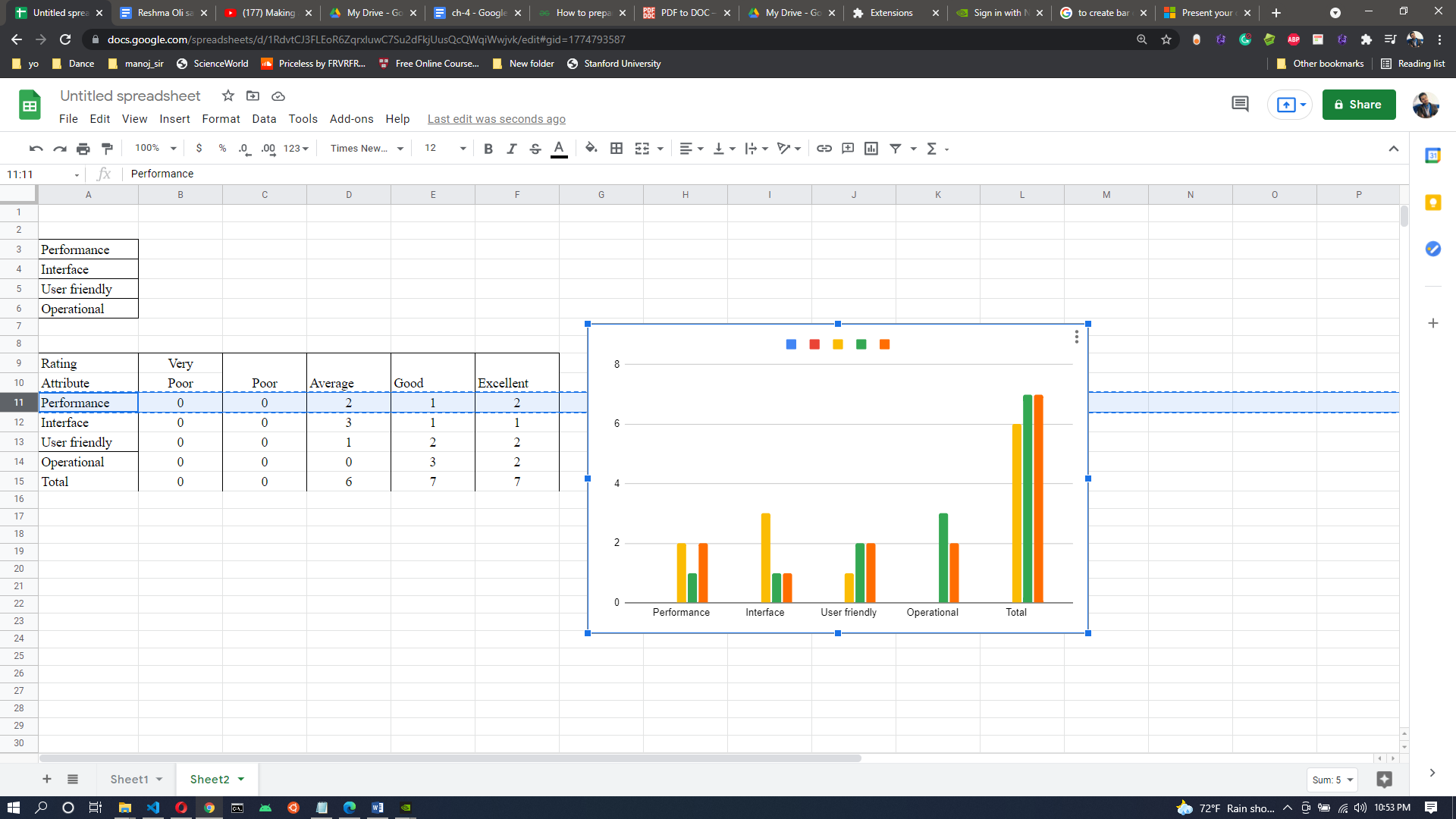
For this testing, five users are selected. The feedback of the users is collected. For this testing two domain exports, two chrome browser users, and one IT bachelor student are selected.

### 5.7.1 Process of Acceptance Testing

* Preparing
  + To ensure that the programs operational flow is correct, a decision table is built.
  + The feedback form template is built
* Planning
  + Providing Feedback form to the end-user.
  + For this testing, the time is set from 7:30 PM to 11:45 PM during night time.
* Opening
  + Introduction
    - Name: Bibek Gautam
    - Programme: Bachelor of Computer Science (Hons.)
    - Purpose: To develop a web-based chrome extension application for detecting fake news articles.

### 5.7.2 Analysis of Gathered Information





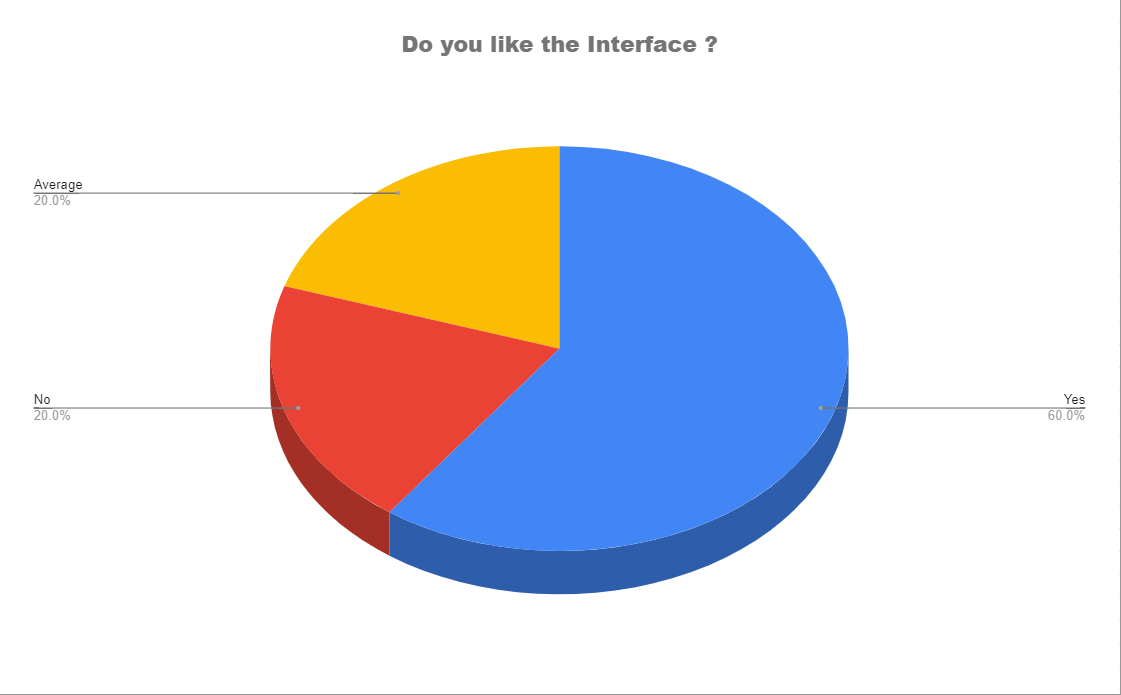


Figure 5.8: Analysis of Gathered information

### 5.7.3 Summary of Gathered Information

Table 5.4: User Rating

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Rating** | **Very Poor** | **Poor** | **Average** | **Good** | **Excellent** | **Percentage** | | |
| **Attribute** | Average | Good | Excellent |
| **User friendly** | 0 | 0 | 0 | 1 | 4 | 0% | 20% | 80% |
| **Interface** | 0 | 0 | 0 | 4 | 1 | 0% | 80% | 20% |
| **Performance** | 0 | 0 | 1 | 2 | 2 | 20% | 40% | 40% |
| **Operational** | 0 | 0 | 0 | 3 | 2 | 0% | 60% | 40% |
| **Total** | 0 | 0 | 1 | 10 | 9 |  | | |
| **Percentage** | 0 | 0 | 5% | 50% | 45% |

From the table dissected data 50% respondent rated for Good, 45% respondent's rate for Excellent and 5% respondent's rate for average. While 20 % of clients like the interface though, 80% of clients appraise for good interface design.

## 5.8 Discussion

In this project, different ML models have been trained and tested. Since the LinearSVC model has the highest accuracy on the given dataset among the other ML models with approx. 96%. LinearSVC model is selected to use for detecting the fakeness of the news article. LinearSVC is well-known for its nonlinear input space kernel technique. Which offers very high accuracy compared to other classifiers. An input data space is transformed into the appropriate form using a kernel. As it creates a hyper plane that draws right in between the negative and the positive result, for now, the chrome extension is connected to the backend part i.e. the detection model and cross-referencing feature. With the help of API. Which is hosted locally. In the future, the API will be hosted in a cloud server. The user input text is submitted to the backend by using the GET method request. The result of the model with the cross-reference of that article is received by the frontend part in the form of JSON. The cross-reference features were accomplished by using Google’s custom search API. The trustworthy sites such as BBC, C-SPAN, The Economist, Kanpur, etc. The uniqueness of this project is its cross-referencing features. End-users can reference the same or similar article from the trusted sites.

## 5.9 Conclusion

In conclusion, the application has been thoroughly tested, four testing methods are utilized to test the application. First, the application's decision is tested by creating a testing table for the application. Furthermore, the application is tried on the integration part to incorporate the ML model. The third strategy is utilized to test the client acceptance and interface testing to gather the application review and comment from the clients. At last, the procedure is utilized to analyze the client rating view.

# 

# 

# **CHAPTER 6**

# **CONCLUSION & SUGGESTIONS**

## 6.1 Introduction

A system requires support and evolution stage to make the system a lot better. Each system consists of bugs, simply there isn't system that does not contain bugs, so to keep up with the security, and the application requires the evolution stage. The application might have mistaken, flaws, and system failures during testing. The correction of the problem can improve the application's quality and reduce risk in the future. The testing is done to lessen the application error and handle the system failure. Thus, this application has met and filled the stated objective and requirement.

## 6.2 Challenges

During the development process, there occurred a few challenges and risks.

* Documentation process Lack of expertise in English to write the documents grammar free. The solution is to use third part application to check the grammatically error in the document such as grammar checker.
* Dataset is the fundamental and most important aspect in training ML modules, obtaining a trustworthy dataset is hard.
* While integrating the POST request to fetch API data was not working

Solution: GET request was used.

* Due to a lack of knowledge, the built extension is not available in the chrome web store.

## 6.3 Advantages of the Application

* Enhanced the User Experience.
* Free license: The extension is free to use without any charges
* It is simple to use, secure and fast.
* Web pages can also be accessed and loaded very easily, even with heavy graphics, advertisements, and video material browsing across several pages.
* It provides automated changes that are enforced to ensure that safety is reviewed.
* It implements automated adjustments to guarantee that safety is monitored.
* Does not have to open a new tab or search the app one tab and the system is ready to go.

## 6.4 Disadvantages of the Application

* Need internet connection to detect fake news.
* No online API is available. Therefore, the application will not operate on other devices.
* The application is only available in chrome web store.

## 6.5 Future Development Cross-Platform

* The extension will be available in different browsers.
* Available in the web store.
* More emphasis on user interface design and development
* The ML model will be optimize by tuning the hyper-parameters.
* New dataset will be tested to obtain more accuracy.
* Complete analysis of the model to improve its performance.

## 6.6 Conclusion

This extension will allow the user to detect fake news through the chrome browser. It will enhance user efficiency and effectiveness. The application user interface and overall performance will be enhanced in the future. Overall, the chrome extension meets the proposed objective of the project.

# **REFERENCES**

Ahmad, I., Yousaf, M., Yousaf, S., & Ahmad, M. O. (2020). Fake News Detection Using Machine Learning Ensemble Methods. Complexity, 2020, 1–11. <https://doi.org/10.1155/2020/8885861>

David Gunning. 2017. Explainable artificial intelligence (xai). Defense Advanced Research Projects Agency (DARPA), nd Web (2017)

Victoria L Rubin, Yimin Chen, and Niall J Conroy. 2015. Deception detection for

news: three types of fakes. In Proceedings of the 78th ASIS&T Annual Meeting. 83.

Rohit Kumar Kaliyar, "Fake News Detection Using A Deep Neural Network", IEEE 2018.

Wenlin Han and Varshil Mehta, "Fake News Detection in Social Networks Using Machine Learning and Deep Learning: Performance Evaluation", IEEE 2019

Ranojoy Barua, Rajdeep Maity, Dipankar Minj, Taranag Barua and Ashish Kumar Layek, "F-NAD: An Application for Fake News Article Detection

Sahil Gaonkar, Sachin Itagi and Rhetiqe Chalippatt, "Detection Of Online Fake News: A Survey", IEEE 2019.

Syed Ishfaq Manzoor and Dr Jimmy Singla, Nikita, "Fake News Detection Using Machine Learning approaches: A systematic Review", IEEE 2019.

Anjali Jain, Avinash Shakya, Harsh Khatter and Amit Kumar Gupta, "A Smart System For Fake News Detection Using Machine Learning", IEEE 2019.

Sahil Gaonkar, Sachin Itagi and Rhetiqe Chalippatt, "Detection Of Online Fake News: A Survey", IEEE 2019.

Karishnu Poddar, Geraldine Bessie Amali D and Umadevi K S, "Comparison of Various Machine Learning Models for Accurate Detection of Fake News", IEEE 2019.

Smitha. N and Bharath. R, "Performance Comparison of Machine Learning Classifiers for Fake News Detection", IEEE 2020.

Rahul R Mandical, Mamatha N, Shivakumar N, Monica R and Krishna AN, "Identification of Fake News Using Machine Learning", IEEE 2020.

Vanya Tiwari, Ruth G.Lennon and Thomas Dowling, "Fake News Detection using Machine learning Algorithms", IEEE 2020.

Fake news detection in online social media Problem Statement. (n.d.). https://www.leadingindia.ai/downloads/projects/SMA/sma\_9.pdf

Zhou, X., & Zafarani, R. (2018). Fake News: A Survey of Research. Detection Methods, and Opportunities. ACM Comput. Surv, 1(1). https://arxiv.org/pdf/1812.00315.pdf

Ermakova, L., Hosni Seffih, Firsov, A., & Gaël Guibon. (2019, April 8). Multilingual Fake News Detection with Satire on Vaccination Topic. ResearchGate; unknown.

https://www.researchgate.net/publication/332803834\_Multilingual\_Fake\_News\_Detection\_with\_Satire\_on\_Vaccination\_Topic

FakerFact: Fake News Detection. (n.d.). Chrome.google.com. https://chrome.google.com/webstore/detail/fakerfact-fake-news-detec/hmcmekfmgfmilmmnicpmkfkccgnfegef?hl=en

Fake News Detector. (n.d.). Chrome.google.com. https://chrome.google.com/webstore/detail/fake-news-detector/alomdfnfpbaagehmdokilpbjcjhacabk?hl=en

Wikipedia Contributors. (2021, January 28). Detecting fake news online. Wikipedia; Wikimedia Foundation. https:// en.wikipedia.org/wiki /Detecting\_f ake\_news\_online

Gaël Guibon, Ermakova, L., Hosni Seffih, Firsov, A., & Guillaume Le Noé-Bienvenu. (2019). Multilingual Fake News Detection with Satire. Archives-Ouvertes.fr. https:// doi.org/https:// halshs.archives-ouvertes.fr /halshs-02 391141

‌FakerFact:Fake News Detection. (n.d.). Chrome.google.com. https:// chrome. google.com/webstore/detail/fakerfact-fake-news-detec/hmcmekfmgfmilmmnicpmkfkccgnfegef?hl=en

Fake News Detector. (n.d.). Chrome.google.com. [https:// chrome.google.com/ webstore/detail/fake-news-detector/alomdfnfpbaagehmdokilpbjcjhacabk?hl=en](https://chrome.google.com/webstore/detail/fake-news-detector/alomdfnfpbaagehmdokilpbjcjhacabk?hl=en)

Debunker - Detect Fake News. (n.d.). Chrome.google.com. [https:// chrome.google .com/webstore/detail/debunker-detect-fake-news/nlmadjgafbedgenmekianiebhdhoccon](https://chrome.google.com/webstore/detail/debunker-detect-fake-news/nlmadjgafbedgenmekianiebhdhoccon)

No Bias News. (n.d.). Chrome.google.com. https:// chrome.google.com/webstore/ detail/no-bias-news/hbbfimdipnhmmipnkngmmofnmenkldio

120 Lies: Fact-Check Amalgamator. (n.d.). Chrome.google.com. Retrieved March 9, 2021, from https:// chrome.google.com/webstore/detail/ 120-lies-fact-check-a malg/jkhkjbehpofkeoligeeepllnbikenimh

‌NewsCracker. (n.d.). Chrome.google.com. Retrieved March 9, 2021 from

https://chrome.google.com/webstore/detail/newscracker/lmpfanpnpoaegbafkodbifallmfcncpb

‌

**APPENDICES B: USER RESPONSE DATA ANALYSIS**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Education Background | How do you rate ML prediction response? | How Do you rate the Interface design? | How do you rate the extension performance? | How do you rate the user friendly? |
| IT Bachelor | Good | Good | Excellent | Excellent |
| Chrome browser user | Excellent | Good | Good | Excellent |
| Chrome browser user | Excellent | Excellent | Excellent | Excellent |
| Domain export | Good | Good | Average | Good |
| Domain export | Good | Good | Good | Excellent |

\*All responses were obtained via online call.