

# 1. Introduction

- Fake news is considered one of the main threats to the global information revolution in the world. They tend to cause serious damages to the reputation of persons or organizations, they also tend to cause many indirect effects on other aspects depending on the type of fake news.
- Fake news elimination is a critical issue in society today, as fast-paced networks have a way of spreading false information and indirectly causing chaos in the world, as is. Existing models for fake news detection have been focused on creating an environment and integrating it with social media. Our model is designed specifically for the application called WhatsApp.
- Whatsapp has been a major source for fake news spreading through nations, creating political indifferences, and causing riots based on sheer fake information spread on WhatsApp, during the covid period, WhatsApp also became the major source of fake data spread through groups, and fake remedies.

## 1.1 Need of the Project

- The prevalent age of social media is fragile towards the sector of news that is accurate. It has been proven, multiple times, that social media has been a messenger for fake news and has created chaos among those who fall for it.
- The crowd on social media is indefinitely naive and has a chance of falling for these traps, our application is targeted at those crowds that blindly follow the news that is forwarded or circulated without holding any accountability or information on what's the truth.
- The model has targeted the audience of those who have a chance for following in for these traps and has specifically enhanced the model for WhatsApp, the future enhancements are incremental and have exponential growth in it.

## **1.2 Approach**

- We approached the problem statement by identifying that the number of fake news that was being spread was through WhatsApp. We created a machine learning model that integrated the dataset of the entire world's news, without bias, and focused majorly on Indian news traditionally.
- We understood that the problem could be solved by implementing a machine learning algorithm that detected the fake news by typing in the box, verifying through the dataset, and presenting if the news was fake.
- The dataset that we have used is from the year 2015-2016 and consists of news from those years only, we used the previous years dataset to increase accuracy and efficiency.

## **1.3 Benefit**

- False information leads to conflicts in our community that will disrupt the harmony of the world, fake information has its way of spreading through media and disrupting economic incentives, because financially motivated, we aim to create a product to curb the spread of false news and help people make more informed decisions when they encounter such false news.

## **1.4 Competition**

- OIGETIT is one of the major competitors with improved accuracy and increased efficiency.
- Twitter's model for fake news detection also has increased datasets included.
- Instagram has also been a major league competitor.

## 2. Customer Validation

### Persona 1:

Persona		Team
<p>I am a Journalist, I have a husband and 2 children and we are living together in Mumbai for past 5 years.</p> <p>About this persona: Palki Kumari Live TV journalist</p>	<b>1 Pains</b> Fears, frustration and anxieties	<b>2 Gains</b> Wants, needs, hopes and dreams
	<p>I have the job of appearing on tv news channel and report news and sharing fake news by mistake live and on global channel is scaring me.</p>	<p>The model should fast detect and obviously point out fake news articles related to any particular source I search for and classify it as a doubtful source of news as fast as possible.</p>
	<b>3 Jobs to be done</b> What are they trying to do and why is it important for them?	<b>4 Reality</b> How do they achieve those goals today? Any barriers in their way?
	<p>Giving false news on global channels can incite riots and hatred among communities or in general and misguide people and they are trying to prevent that.</p>	<p>It can be done by building a fake news detection model that will be simple, fast and accurate enough to detect the false news and classify it as a fraudulent source of news as quickly as possible to avoid wrong reporting.</p>
<b>5 Stories and observations</b> Write down quotes or observations that best describe their experience		
<p>I was able to escape many fraudulent news sites using this model. I was alerted many times when I chanced upon a site that would spread false information and try to steal data and prevent me from mistakenly reporting some news that might be fake or wrong on global channels.</p>		
<b>6 Context</b> Are there other factors that we should take in consideration?		
<p>There are factors like not all news is fake, and it might be probable that some news was not fake but just originated from a shady background. There must be some way to report such sites with frequent patterns of spreading false news and misguiding people.</p>		

## Persona 2:

### Persona

Team

I am a news report writer in Delhi, I have a wife and we are living together in Ghaziabad for past 3 years.

#### About this persona:

Kailash sharma

News report writer

#### 1 Pains

Fears, frustration and anxieties

Worried I would end up spending too writing wrong journals or articles on fake news and get fired from my job.

#### 2 Gains

Wants, needs, hopes and dreams

The model used must be accurate and should find out fake news that circles around and classify it as a doubtful source of news as fast as possible.

#### 3 Jobs to be done

What are they trying to do and why is it important for them?

---

---

---

---

#### 4 Reality

How do they achieve those goals today? Any barriers in their way?

---

---

---

---

They are trying to write articles, reports and journals on news for customers who are invested in news and they are trying to write only on correct data and information.

By referring to the model that will be simple and fast enough to detect the false news and classify it as a fraudulent source of news as quickly as possible.

#### 5 Stories and observations

Write down quotes or observations that best describe their experience

Using this service, I was able to escape many fraudulent news sites and was alerted many times when I chanced upon a site that would spread false information and try to steal data. It also helped me identify where and which sites and places fake news often originated or came from.

#### 6 Context

Are there other factors that we should take in consideration?

We should also consider that not all fake news has similarities and hence does not have patterns that can be detected easily. Also, many of the news sources are not completely fraudulent.

## Persona 3:

### Persona

Team

<p>I am a business analyst and I work around shares and stock market. I have a daughter and I am a widower.</p> <p><b>About this persona:</b> Ravi Kumar Stock market and data analyst</p>	<p><b>1 Pains</b> Fears, frustration and anxieties</p> <p>I am scared of fake news circling ,about shares and stock market prices being spread and it can mess up my job and my clients connection.</p>	<p><b>2 Gains</b> Wants, needs, hopes and dreams</p> <p>The model should fast detect and obviously point out fake news articles related to any stock/share market source I search for and classify it as a doubtful source of news as fast as possible.</p>
	<p><b>3 Jobs to be done</b> What are they trying to do and why is it important for them?</p> <p>_____</p> <p>_____</p> <p>_____</p> <p>_____</p>	<p><b>4 Reality</b> How do they achieve those goals today? Any barriers in their way?</p> <p>_____</p> <p>_____</p> <p>_____</p> <p>_____</p>
	<p>They are trying to analyse right data and has work that involves share market news and business news as well stock market analysis and they need to have right data for such a job.</p>	<p>It can be done by building a fake news detection model that will be simple, fast and accurate enough to detect the false news and have categories so that it can look up false business related news and classify all doubtful sources.</p>

**5 Stories and observations**  
Write down quotes or observations that best describe their experience

I was able to sort down all malicious sites reporting false news on business and market, and I was able to compare and analyse data across different sites and verify the accurate information.

**6 Context**  
Are there other factors that we should take in consideration?

Some sites give wrong numbers, which can mess up critical information on business and the market. People need to know where they are investing and if they are investing or considering right information for their own sake. These sites should be reported or listed for risk and prevention.

## Persona 4:

### Persona

Team

I am a normal housewife who lives with her husband in Noida, India with her two sons.

#### About this persona:

Meena Kumari  
Housewife

#### 1 Pains

Fears, frustration and anxieties

I am scared of fake news circling related to coronavirus and other diseases which scam people on an everyday basis.

#### 2 Gains

Wants, needs, hopes and dreams

The model should fast detect and obviously point out fake news, frauds etc that are often forwarded on social media platforms like whatsapp etc.

#### 3 Jobs to be done

What are they trying to do and why is it important for them?

---

---

---

---

---

#### 4 Reality

How do they achieve those goals today? Any barriers in their way?

---

---

---

---

---

They are trying not to get scammed by whatsapp forwards that spread false news and can incite wrong information about anything ranging from medicine to wrong medication.

It can be done by building a fake news detection model that will be simple, fast and accurate enough to detect the whatsapp forwards and check their validation and save people from frauds and protect users' privacy.

#### 5 Stories and observations

Write down quotes or observations that best describe their experience

The model helped a lot regarding fake news such as coronavirus medication and do's and don'ts circulating on Whatsapp and other social media platforms and thus unnecessarily hindering people's general knowledge.

#### 6 Context

Are there other factors that we should take in consideration?

People need to know the legitimization of news being forwarded on WhatsApp and other social media platforms and prevent themselves from being scammed by false medication news or scare news which are meant to harass people mentally as well as try to steal the data of users through such means. The ways to detect the sources of such forwards can be improvised as a result.

board of  
innovation

**About this tool**  
Personas are fictional profiles, often developed as a way of representing a particular group based on their shared interests and needs.

Personas can provide a range of different perspectives, allowing teams to define and engage the different target groups that may exist within their target market.

©2020  
personas AI  
boardofinnovation.com/boards

## Persona 5:

### Persona

Team

I am an aspiring JEE student who lives in Gurgaon with my parents for a decade now.

#### About this persona:

Meena Kumari  
Housewife

#### 1 Pains

Fears, frustration and anxieties

Used to be scared of fake news regarding JEE exam during the coronavirus. There are wrong news about results and exam centres which harasses students mentally.

#### 2 Gains

Wants, needs, hopes and dreams

The model should detect coronavirus related fake news as well as news that include JEE/NEET and other exams.

#### 3 Jobs to be done

What are they trying to do and why is it important for them?

---

---

---

---

---

#### 4 Reality

How do they achieve those goals today? Any barriers in their way?

---

---

---

---

---

They are trying not to get scammed by whatsapp forwards as well on other sites like Twitter that spread false news and can incite wrong information about examination during coronavirus or in general regarding JEE/NEET and other exams.

The model should be built to detect whatsapp forwards especially and fake news on social media platforms that can jeopardise a student's mental health and spread false news about results and exam centres and allotments.

#### 5 Stories and observations

Write down quotes or observations that best describe their experience

The model helped a lot regarding fake news, such as detecting fake exam centres' news and allotments of seats and results that might have misguided a student's decision-making process or incited a false assumption.

#### 6 Context

Are there other factors that we should take in consideration?

People need to know the legitimization of news being forwarded on WhatsApp and other social media platforms and prevent themselves from being scammed by false examination news being spread on Twitter and WhatsApp which are meant to harass student mentally and jeopardise students' futures by misguiding them.

board of  
innovation

About this tool  
Personas are fictional profiles, often developed as a way of representing a particular group based on their shared interests and needs.

Personas can provide a range of different perspectives, allowing teams to define and engage the different interest groups that may exist within their target market.

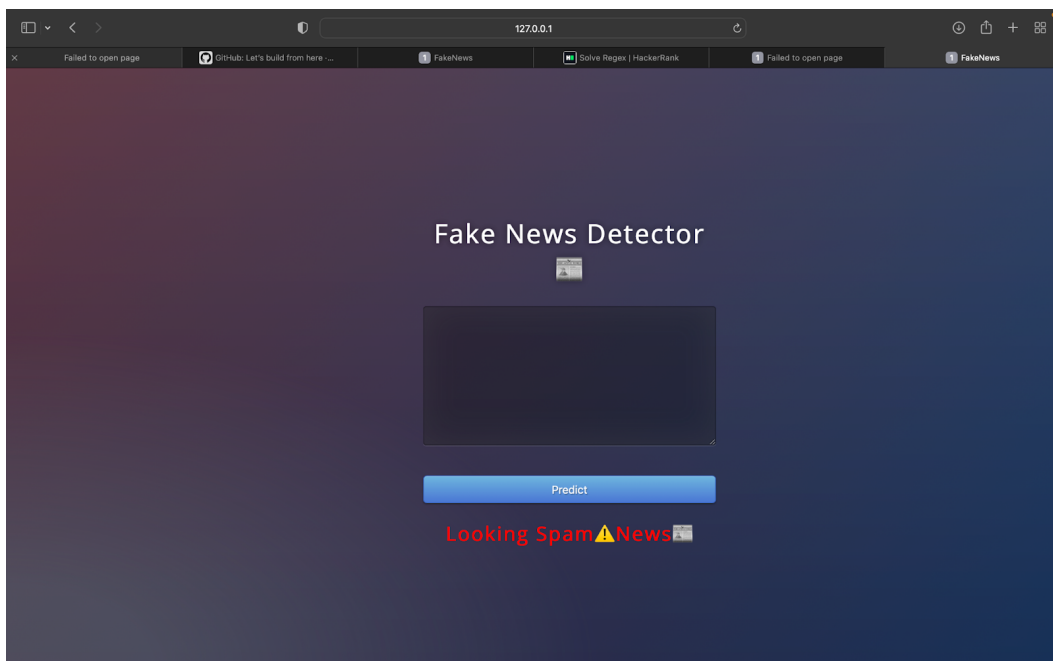
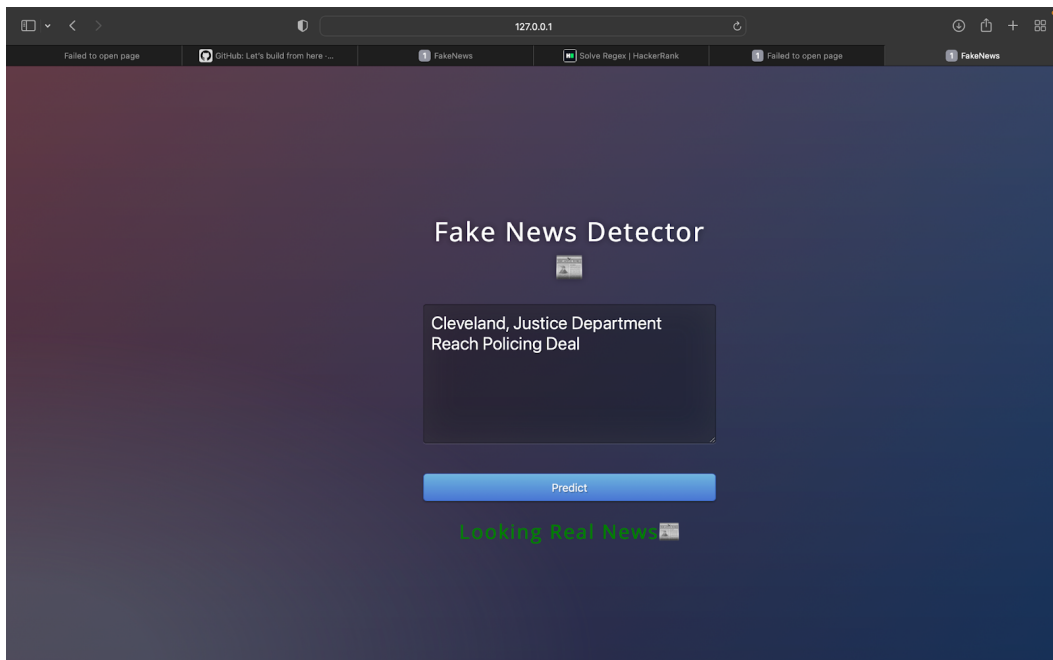
©2020  
personae A1  
boardofinnovation.com/boards

### 3. Project Description

- In the recent times a lot of fake information has been virally spreading through social
- media platforms which has created a lot of political tension along with fake information
- being rapidly spread and has resulted in a lot of chaos around the globe.
- This has led to a lot of conflict among people regarding beliefs and trust including action.
- People being vulnerable tend to believe the fake information without cross checking it
- with actual sources and further go on to spread the information
- Now this problem can be tackled if there is a system to identify or regulate information
- that is being passed on by untrusted sources
- During the covid peak , a lot of fake information was being flooded on social media and
- people were not able to differentiate if it was fake or not. Which caused a lot of unrelated data and information flow.
- The solution to tackle this problem can be building a machine learning model that filters through information and data from actual sources.
- Recently Twitter has added this feature where the app can cross check information and tell if it's fake or reliable.
- The model will help detect and analyze the fake information and tell if its reliable or
- not. This way the fake information being passed along with people spreading it will be in control.
- The model will help avoid conflicts and disputes along communities, people and nations.
- Datasets have been created especially for this project in addition to the online available data sources. The sentiment analysis model has been done using the deep learning model and it has achieved 87% accuracy, while the objectivity check has not achieved significant results. News sources analysis problem has been dealt with as a traditional term frequency problem, the solution achieved a 94% accuracy value. Due to the lack of enough data sources, the fact-checking solution has ended up in creating a dataset that is ready for fact-checking against any relational dataset of periodic values such as the stock market.



### 3.1 Illustrate the UI/ Input, output:



### 3.2 Technical Components of the project :

Pandas - for data manipulation and dataframe

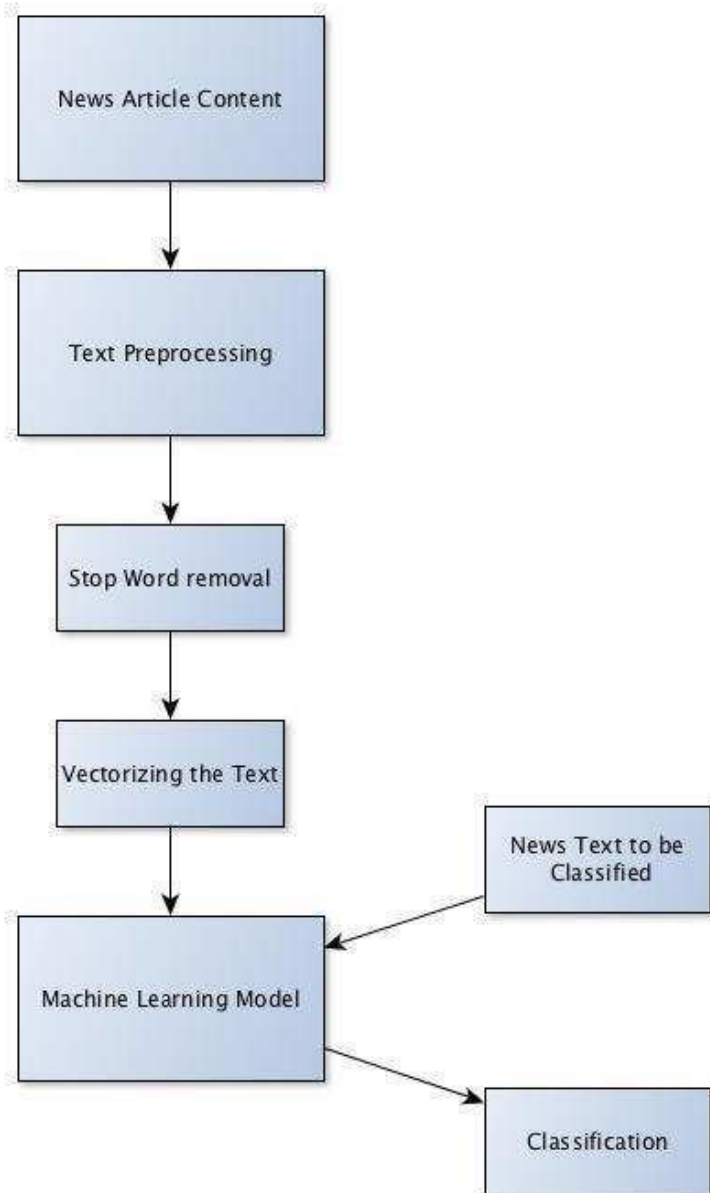
- Scikit-learn - TfidfVectorizer
- Passive Aggressive Classifier : A machine learning algorithm to train by feeding it instances sequentially in mini batches
- Pickle - to export the trained model

- We look forward to implement the same model in BERT Model basically BERT stands for Bidirectional Encoder Representations from Transformers

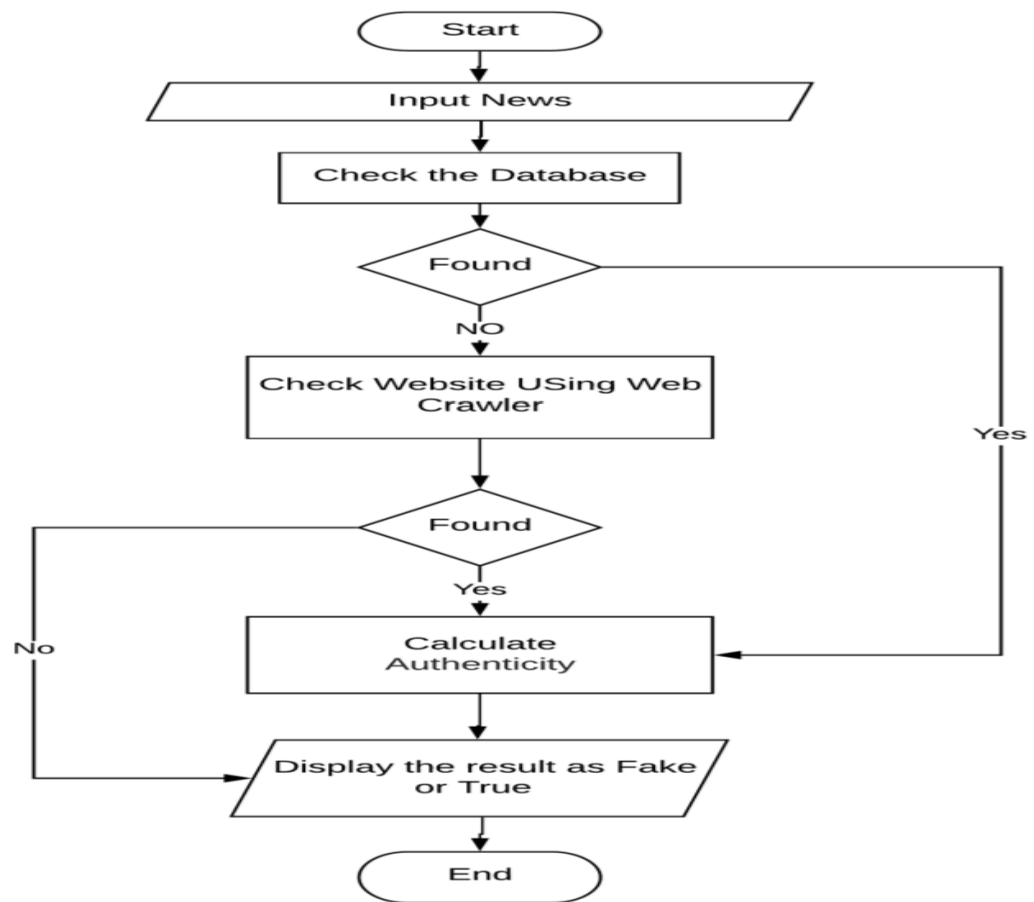
- BERT's key technical innovation is applying the bidirectional training of Transformer, a popular attention model, to language modeling.

- This is in contrast to previous efforts which looked at a text sequence either from left to right or combined left-to-right and right-to-left training.

### 3.3 System Architecture:



### 3.4 Data Flow in the system:



**Fig. 2: Flow Chart.**

## 4. Business Plan:

A business plan to tackle the above problem statement has been derived and has identified its main business objectives, summaries, finances, market research and strategy with valid documentation has been devised and stated below.

- Business Objective: The objective of this business is to tackle the need for false information and provide a model that has been trained to determine if the news is relevant or not. The audience stays more informed and can due to this actively participate in decision making without bias.
- Summary: The model has been devised to reduce the percentage of fake and increase information towards the citizen.

### 4.1 Key activities

The key activities for Business Plan are :-

#### **Marketing**

The software would be advertised to all social media users as a way to ensure that they get real news and don't get scammed or misinformed by fake forwards on social media.

**Sales Promotion(For New Customers)**

There would period when software will have temporary offer which would increase demand for it.This way we would retain customers and attract more customers

**Review from Customers**

We will take surveys from customers periodically.We will ask whether the software was accurate and what they are expecting from software.

**Enhancement**

Based on reviews we would enhance the algorithm so that the software becomes more accurate.

## **4.2 Key Resources**

The key resources are:-

### **Software Development**

The main software ,that detects whether news from social media is fake or real, is the main resource of the business. Development of software is the main priority.

### **Experts**

Experts on development of software and social media to help us with development of algorithms.

### **New Datasets**

New data set should be acquired and used to train the algorithm so that the algorithm is up to date.

### **Improving Algorithm and Analysis**

Algorithm can be improved on the basis of reviews and its analysis can be improved by using new datasets to train the algorithm.

### **Updatations**

Periodic patches for software will be released to fix the bugs and problems that the customers face, updating the dataset according to the yearly newsstand, increasing the accuracy of the

## 4.3 Key Partners

The key partners for business are:-

### **Social media platforms**

Big social media platforms would be our partners as they would like to use our software to make their social media more free of fake news. This would attract customers to their platform.

### **Investors**

Investors would invest in the business in return for a percentage of profit.

### **Journalists**

Journalists can use our software to verify if news is real or fake. This would further advertise our software.

## 4.4 Value Propositions

### **News credibility**

Main purpose of our software is to check the credibility of news. It is to tell whether the news is real or fake. It is to inform the user of the credibility of news.

### **Fake news spreading less**

Our purpose is to reduce the spread of fake news and inform users of it.

### **Community fights with fake news can be prevented**

Misconception can be prevented from reducing the spread of fake news and would help people bond and not fight.

### **Scams can be prevented**

People will also get less scammed as software would verify for them whether it is fake or real.



## 4.5 Cost Structure

### **Software development costs**

Cost incurred during development of software and buying equipment related software.

### **Legal requirements costs**

Legal cost for software copyright, patent, terms and conditions, payment for establishing, etc.

### **Accurate news check**

Investment in fact checking mechanisms which can be automated and manual.

### **System installation checks**

### **Cloud**

Cost incurred for SaaS and PaaS.

### **Research and Development costs**

Cost incurred when researching for data and data set for the fake news detection software.

### **Model training equipment**

Cost incurred for equipment training algorithms with data sets.

## 4.6 Revenue Streams

### **Installation fees**

The fees paid by customer when buying our software services

### **Third party software updates**

The revenue earned during updation by third party software.

### **Marketing and increased accuracy**

### **Upgradation costs**

Fees paid by customers for getting an upgraded version of our software.

## 4.7 Customer Segment

### **B2B based business**

### **IM platforms**

Third party launch for Instant Messaging platforms to prevent small scale spread of fake news.

### **Social Media platforms**

Take over social media as a bot account to flag tweets from news providers as fake or real.

## 4.8 Customer Relationship

### **Feedback Analytics**

Use feedback and customer flags to reclassify tags.

### **Observational Analytics**

Base further updates on usage patterns to optimize experience for businesses.

## 4.9 Channels

Service will be provided only via established platforms.

B2C channels will not be used.

## **5. Financial Plan:**

### **5.1 Growth Strategy**

#### **Reinforce service values and product proposition**

A strong service will be more attractive to platforms.

#### **Focus on UI/UX**

The platforms will want a simpler UI/UX for their users.

#### **Inbound Marketing**

Start with free services and capture a target audience.

Gradually funnel down into business.

#### **Allow platforms for customization**

Platforms would want the service to be implemented as smoothly as possible for their users. For this they would require to make several tweaks in the same.

### **5.2 Traction**

Initially focus on generating awareness among key influencers.

Creating awareness among a group for beta testers.

Run a loyalty program for the platforms using our service.

## **5.3 Financials**

### **Heads of income**

- Digital drives
- Customer revenue
- Service patent

### **Heads of expenditure**

- Customer Service
- RnD costs
- Data acquisition
- Advertisements
- Miscellaneous

Table of Financials for Three Years

	2022	2023	2024	
<b>Users</b>	52,000	401,000	1,670,000	
<b>Jobs</b>	509,000	4,005,000	16,009,000	
<b>Average price per job</b>	75	80	90	
<b>COMPANY REVENUE @15%</b>	5,625,000	48,000,000	216,000,000	
<b>- Cost of Revenue</b>	0	0	0	
<b>Gross Profit</b>	5,600,000	48,001,000	236,000,000	
<b>OPEX</b>				
<b>- Sales &amp; Marketing</b>	5,062,500	38,400,000	151,200,000	70%
<b>- Customer Service</b>	1,687,500	9,600,000	21,600,000	10%
<b>- Product Development</b>	562,500	2,400,000	10,800,000	5%
<b>- Misc.</b>	281,250	2,400,000	4,320,000	2%
<b>TOTAL OPEX</b>	7,596,750	52,800,000	187,920,000	
<b>EBIT</b>	<b>-1,850,750</b>	<b>-4,500,000</b>	<b>28,080,000</b>	13%

## **6. Conclusion :**

The fake news detection using machine learning and dataset training has been concluded by defining objectives, testing as well as validation has been accomplished.