# AN INTELLIGENT NEWS-BASED STOCK PRICING PREDICTION USING AI AND NATURAL LANGUAGE PROCESSING

Sirui Liu[1] and Yu Sun[2]

[1]Orange Lutheran High School, 2222 N Santiago Blvd, Orange, CA 92867
[2]California State Polytechnic University, Pomona,
CA, 91768, Irvine, CA 92620

***ABSTRACT***

*How do you know which stock is the right stock to invest in and have no risk of losing their money [1]? Even though there are analysis specialists out there to collect data to calculate which stock is good to be invested in, ultimately people could not afford the cost of specialists and specialists are not able to be there every minute that you want to find them. Therefore, the app Stock Recommendation is created to solve this problem where stock investment suggestions are available in touch anywhere and anytime [2]. This application helps us with what we want to invest in and gather information from recent news to show us about the public opinions towards the stock that we are looking for. Investors will no longer struggle with the problem that is the stock that they want to invest in, a good stock or a bad stock, so no money will be lost from the investor's pocket and rather, they will gain my money [4].*

***KEYWORDS***

*Stock, machine learning, AI.*

## 1. INTRODUCTION

The Great Depression led to the first time that the stock market officially went into everyone's sight with how much the stock market changed society and people's life [9].

statistic #1: On any given day, stocks have roughly a 53 percent chance of rising and a 47 percent chance of falling. Over any given 3-month period, stocks rise 68 percent of the time, dropping the other 32 percent of the time.

statistic #2: A year after the Covid pandemic shut down the economy, stocks have gained 79% from the lows and the market is in a solid position to continue to rally. It's now being led by sectors that had been very unlikely leaders — like energy and industrials.

Some of the people had discovered and studied multiple ways to predict and proposed how the stock system will go while calculating based on the articles and statics that websites like CNN and Yahoo gives on News that allowed the stock buyer to get an understanding of which stock they should be invested in bases on the news articles that they have published about the recent stock movement that was happening [3]. However, a huge percentage of people who buy stocks do not go to a professional stock adviser but instead, they watch news to see the numbers of the stock. Their sources of information are very limited in the limited source of information that they observed from, with samples given of CNN and Yahoo News being mostly the only two sources

David C. Wyld et al. (Eds): CONEDU, CSITA, MLCL, ISPR, NATAP, ARIN - 2022
pp. 147-155, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.121011

that normally people who buy stocks get information from. The limitations of only two sources are there to provide information and create a lot of limitations by the source preferences and their subjective opinions about a stock. Other techniques to calculate for should a person invest in a stock, for example, stock analysts [5]. They not only take a lot of time to analyze and give advice, but also charge a lot of money to do the stock investigation suggestions. However, plenty of time, stocks are not able to be calculated (eg. GameStop), and often results in losing huge amounts of money with investigation. A second practical problem is that giant amount of stock buyer do not relate the recent activities of one company to its stock, which forms the problem that they might keep in or sell of a stock because it shows that it is losing money or gaining money based on what the curve shows right now, and not looking forward to the future possible incomes.

Yahoo finance and CNN are two resources of stock market movement that a huge amount of people look up to for the purpose of seeking information and carefully think about their investigation towards a stock [6]. Yahoo finance and CNN are both news resources and search engines that provide information about daily stock's curve and statistics about how much a stock increases or decreases.

The app that is being built is an app that shows either positivity or negativity about whether you should invest in a stock or not. The app uses Google API to find, point out, and analyze the positive or negative words that are being found in the news resources linked in the app, and rate how positive and how negative the news is about the stock and the company that you want to invest it in.

The app gives actual suggestions about should you invest in the stock compared to Yahoo finance that just gives you a lot of statistics about the stock but not real suggestions about should you invest in the stock or not. Also the app links to more search engines compared to Yahoo finance that only has its single source of statistics.

In the application of the stock predictor, we will have two ways to demonstrate the usage. First, we show the validity of the prediction results by separating the training set and validation set. By different ways of partitioning data into training set and validation set, we can validate the accuracy of the stock prediction at each given period from the training set. By comparing the result of each training set with the validation set, a validation matrix can be computed. Through the validation matrix we can analyze the potential fitting of the machine learning function. Second, we analyze the usability of the application through a user likeability survey. Different users will try out the application and provide a subjective response based on their interaction experience. It will be analyzed with whether they agree with the trending or not. They will also rate the application based on its aesthetic value.

Introduction of the background, open problem, solution and special contribution, and paper structureThe rest of the paper is organized as follows: Section 2 gives the details on the challenges that we met during the experiment and designing the sample; Section 3 focuses on the details of our solutions corresponding to the challenges that we mentioned in Section 2; Section 4 presents the relevant details about the experiment we did, following by presenting the related work in Section 5. Finally, Section 6 gives the conclusion remarks, as well as pointing out the future work of this project.

## 2. CHALLENGES

In order to build the project, a few challenges have been identified as follows.

### 2.1. Understanding the reports

For most people yahoo finance and CNN data reports are too complicated to understand which lead to a point that they start to buy the "wrong" stock that will make them lose their money in the stock market [7]. By 2018 this dropped 8 percentage points to 34%. More alarming, less than one-third of adults understand three basic financial literacy topics by age 40, although many important financial decisions are made decades earlier. This becomes a very important problem since people start to take risks in the stock market and put plenty of their money into stocks, which they end up buying the wrong stock that makes them lose all of their money.

### 2.2. Choosing a method to enter the stock market

There are too many resources for stock, so it is overwhelming for people to try to enter into the stock market [8]. Thousands of websites, books, magazines, and it can be very overwhelming when people want to find out one single piece of information that they need. Most times too many news articles and news information resources have identical but conflicting sources, which about one thing, each website might have the same information but each website has a different opinion. As a result of causing confusion that makes people not understand and know which stock they should have invested in or which stock is the correct stock to invest in.

### 2.3. Understanding machine learning websites

A lot of websites already use machine learning, but it's complicated and hard to understand. Some already use linear regression, etc [11]. We use sentiment analysis which is very good at classifying whether pages are happy or sad, good or bad news, so it's easy for people to understand how it works and what it does and it makes people feel more comfortable using our app.

## 3. SOLUTION


Figure 1. Stock recommendation

The application has been implemented using python and flutter, and it is carefully developed to serve as a multi-functional platform to support the visually-impaired population in navigation, during natural disasters, and in the midst of the COVID-19 pandemic [13]. The application intends to take all aspects into consideration when it provides features like QR code login, locative marker placement, vibration when detected obstacles, alert in face of disasters, GPS-frequency database, and sanitation reminder [14].

The result shows the company's movements that end up reflecting whether you should invest in or not invest into a company based on their recent news articles. By the calculations that API do, they are able to catch emotion words and rate the emotions inside of the words.

The blue links are clickable to actually browsing the website that the API gets information from since the API is still just robots, people might end up having different feelings towards the same word. So blue links that direct to the actual website are provided to let users read it themselves and think about it if they do not trust the result the app gave.


Figure 2. Screenshot of using page

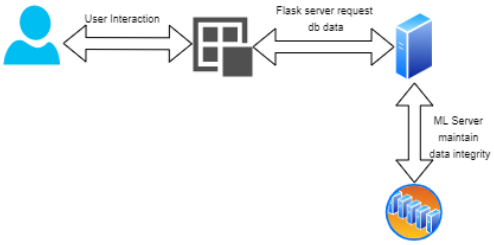


Figure 3. An overview of the project

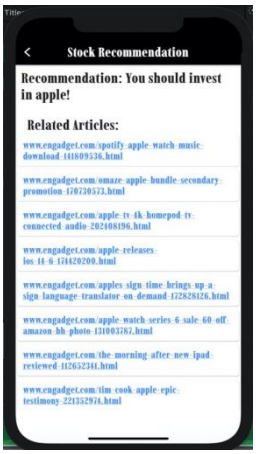


Figure 4. Screenshot of related articles

The application is designed with two main components, the front-end display and the back-end server that provides the data to be displayed. For the front end we used flutter to design the front end through the graphical interface. In the front end a user input box is provided for the user to enter the name of the company for the stock they wish to gain more information with. The user

input will be saved and sent to the server through GET request. The server is written in flask with several API that provide necessary information for the front end such as articles through json.

```
35
36  class MyHomePage extends StatefulWidget {
37    MyHomePage({Key key, this.title}) : super(key: key);
38
39    // This widget is the home page of your application. It is stateful, meaning
40    // that it has a State object (defined below) that contains fields that affect
41    // how it looks.
42
43    // This class is the configuration for the state. It holds the values (in this
44    // case the title) provided by the parent (in this case the App widget) and
45    // used by the build method of the State. Fields in a Widget subclass are
46    // always marked "final".
47
48    final String title;
49
50    @override
51    _MyHomePageState createState() => _MyHomePageState();
52  }
53
54  class _MyHomePageState extends State<MyHomePage> {
55    Widget build(BuildContext context) {
56      Timer(
57        Duration(seconds: 5),
58          () => Navigator.of(context).pushReplacement(
59              MaterialPageRoute(builder: (BuildContext context) => InfoPage()))
60      );
61      return Scaffold(
62        backgroundColor: Colors.black,
63        body: Container(
64          margin: EdgeInsets.fromLTRB(20, 90, 20, 0),
65          child: Column(
66            mainAxisAlignment: MainAxisAlignment.start,
67            crossAxisAlignment: CrossAxisAlignment.center,
68            children: <Widget>[
69              FittedBox(
70                fit: BoxFit.contain,
71                child: Text('Stock\nRecommendation',
72                  style: TextStyle(color: Colors.white, fontSize: 60, fontFamily: 'Imbue', fontWeight: FontWeight.w600),
73                ),
74              ),
75              Padding(
76                padding: EdgeInsets.fromLTRB(0, 40, 0, 40),
77              ),
78              Image(
79                image: AssetImage('assets/images/stockicon.png'),
80                height: 250,
81              ),
82              Padding(
83                padding: EdgeInsets.fromLTRB(0, 40, 0, 40),
84              ),
85            ],
86          ),
87        ),
88      );
```

Figure 5. Screenshot of code 1

The design of the front end is shown above. There are three main components in the front end, which are the main page, info page, and the results page. The main page displays the question to the user to ask for the company they wish for more information about. After the user enters the information, it will parse the information, send to the information page then redirect to the request to the server (Served at https://Stock-thing.oxxxm.repl.co/results/$company). The API will then return a json list to the result page, where the returned information will be split into a list and displayed to the front end.

```
def getArticles(company):
    # Init
    global articles, topic
    topic = company

    newsapi = NewsApiClient(api_key='5ab7a52681914e49813c2ee13f4141e4')

    d = datetime.datetime.strptime(str(date.today()), "%Y-%m-%d")
    d2 = str(d - dateutil.relativedelta.relativedelta(days=7))

    # /v2/everything
    all_articles = newsapi.get_everything(
      q = topic,                          sources = 'ars-technica, business-
insider, the-verge, bloomberg, engadget, fortune, techcrunch, techradar, the-wall-
street-journal, wired',
domains = 'marketwatch.com, fool.com, finance.yahoo.com, morningstar.com,
seekingalpha.com, investopedia.com, zacks.com,aaii.com, barrons.com,
kiplinger.com, cnbc.com, thestreet.com',
from_param = str(date.today()),              to = d2[:10],
language ='en',                          sort_by ='popularity',
page_size = 100,                         page = 1)

    articles = all_articles['articles']

```

Figure 6. Screenshot of code 2

The servers are written in Python using the flask library, where the application is created using app=Flask(app) command. The server hosts one API which is retrieveJson. In its parameter a company is entered which is passed from the front end. Based on this company name it calls the getArticles(Company) function where we use news api to search for related information. Once the results are fetched from the api, the information will be saved in a global variable called links. The function named getScored will then be used to parse each link's information semantically and generate a score for each returned article. If the article has a score between -15 to 15, the article will be used and displayed at the front end. Lastly, once this information is correctly scored and a result list has been finalized, the toJson function will jsonify the results package into json format, and the json results will be returned to the front end.

## 4. EXPERIMENT

### 4.1. Experiment 1

A good user experience is as important as a good product. So a perfect solution should have excellent user experience feedback. In order to prove that our solution has the best user feedback, we specially designed a user experience questionnaire base on the US system usability questionnaire rules We statistics the feedback result from 100 users, Show the user our app for 1-5 minutes, let them explore freely on the functionality. We divide those users into Five different groups. The first group of users ages from 10 - 20, the second group of users ages from 20 - 30, the third group of users ages from 30 - 40, the fourth group of users ages from 40 - 50, the fifth group of users ages from 50 - 60The goal of the first experiment is to verify high feedback scores shows high performance We collect the feedback scores form these 5 different group of users and analyze it. Experiments have shown that users who ages from 30 - 40 give the highest result feedback to our app. Which may because of the age between those range are more likely to put their money in stock market [10]. The experiment graph shows below:

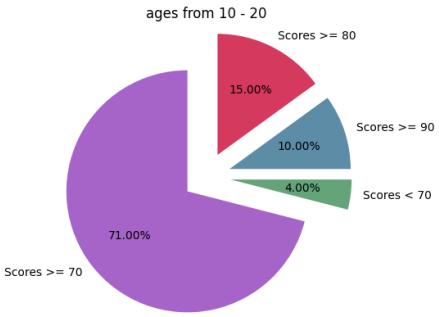


Figure 11. Results of age 10-20

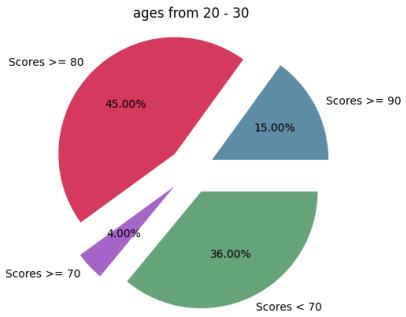


Figure 12. Results of age 20-30

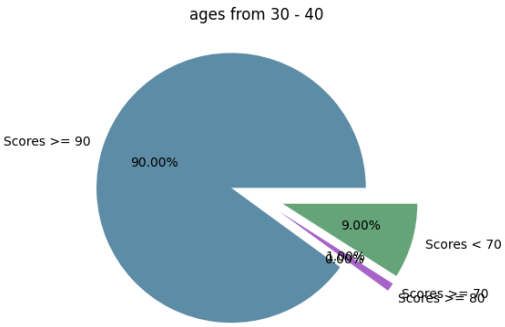


Figure 13. Results of age 30-40

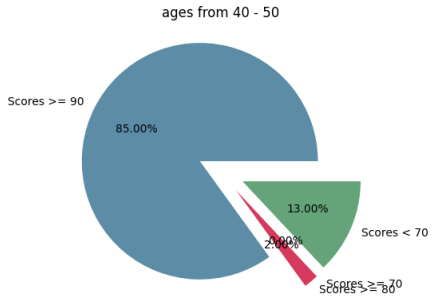


Figure 14. Results of age 40-50

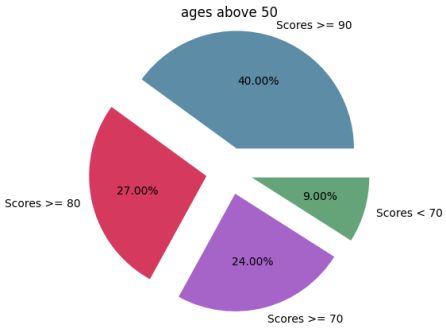


Figure 15. Results of age above 50

## 5. RELATED WORK

The main contribution is that the data and the api reflect the recent news of a company, and show investors that if the company was positive or negative on the recent news reports and websites.

The application was done by using repl.it and flutter which repl.it was used to write down the code to run the program of the app and the flutter was used to build the app. For example, the font, the color and different pages of the app were built in flutter. Android Studios was used to test out the app on the phone to show the errors and what needs to be improved.

## 6. CONCLUSIONS

In this paper, we propose a stock trading information collection system to help stock traders to acquire the information they need. We designed a user-end application using flutter,

- Propose a method/an application

In this paper, we proposed a stock information collection system based on a flutter platform using machine learning and front-back end development [12]. Through this application users will be able to enter the name of the company they wish to find more related information and the application will search through the yahoo database then display the results [15].

- Apply the method/application to experiment

The application was then tested through a usability test with participants number of xxx. Each of the users used the application for five minutes and rated the application through a systematic usability test survey.

- Experiment results indicate its effectiveness and solve challenges

The result indicated that the overall usability is above average according to the usability organization, with a score of xx it indicates that the application has an above average usability score and is easy to adapt as a system.

The application has several limitations. First it only allows you to search the keywords for the company only. Yet the application does not have ways to provide more interactive searching methods.

## REFERENCES

[1] Cutler, David M., James M. Poterba, and Lawrence H. Summers. "What moves stock prices?." (1988).
[2] Walker Z, McMahon DD, Rosenblatt K, Arner T. Beyond Pokémon: Augmented Reality Is a Universal Design for Learning Tool. SAGE Open. October 2017. doi:10.1177/2158244017737815
[3] Gidofalvi, Gyozo, and Charles Elkan. "Using news articles to predict stock price movements." Department of Computer Science and Engineering, University of California, San Diego (2001): 17.
[4] Barber, Brad M., and Terrance Odean. "The behavior of individual investors." Handbook of the Economics of Finance. Vol. 2. Elsevier, 2013. 1533-1570.
[5] Moshirian, Fariborz, David Ng, and Eliza Wu. "The value of stock analysts' recommendations: Evidence from emerging markets." International Review of Financial Analysis 18.1-2 (2009): 74-83.
[6] Xu, Selene Yue, and C. U. Berkely. "Stock price forecasting using information from Yahoo finance and Google trend." UC Brekley (2014).
[7] De Bondt, Werner FM, and Richard Thaler. "Does the stock market overreact?." The Journal of finance 40.3 (1985): 793-805.
[8] Barro, Robert J. "The stock market and investment." The review of financial studies 3.1 (1990): 115-131.
[9] Barsky, Robert B., and J. Bradford De Long. "Why does the stock market fluctuate?." The Quarterly Journal of Economics 108.2 (1993): 291-311.
[10] Aggarwal, Rajesh K., and Guojun Wu. "Stock market manipulations." The Journal of Business 79.4 (2006): 1915-1953.
[11] Su, Xiaogang, Xin Yan, and Chih-Ling Tsai. "Linear regression." Wiley Interdisciplinary Reviews: Computational Statistics 4.3 (2012): 275-294.
[12] Jordan, Michael I., and Tom M. Mitchell. "Machine learning: Trends, perspectives, and prospects." Science 349.6245 (2015): 255-260.
[13] Ciotti, Marco, et al. "The COVID-19 pandemic." Critical reviews in clinical laboratory sciences 57.6 (2020): 365-388.
[14] Tiwari, Sumit. "An introduction to QR code technology." 2016 international conference on information technology (ICIT). IEEE, 2016..
[15] Callery, Anne, and Deb Tracy Proulx. "Yahoo! cataloging the web." Journal of internet cataloging 1.1 (1997): 57-64.