**Extracting Fashion Related News from Facebook**

**Table of contents**

1. [**List of Figures:………..……………………………………………………………………..5**](#_bookmark0)
2. [List of Tables:…………………..……………..…………………………………………… 5](#_bookmark0)
3. ABSTRACT[…………………………………………………………………………………6](#_bookmark1)

CHAPTER 1 ….. [7](#_bookmark2)

1. Introduction[….….………………………………………………………………………………………7](#_bookmark2)
   1. Social Media Impact on Society…..……….…………………………………………..……………8
   2. Fashion Inductry Target through Social Media…………………………………………..,……..12
   3. Background of the Study [….………………………………………………………………………15](#_bookmark3)
   4. Objectives […....……………………………………………………………………………………..16](#_bookmark2)
   5. Research Questions […..……………………………………………………………………………16](#_bookmark3)
   6. Ethical Considerations [….…………………………………………………………………………16](#_bookmark4)
   7. Project Timeline […...……………………………………………………………………………….17](#_bookmark5)

CHAPTER 2 .. [………………………………………………………………………………………...19](#_bookmark2)

1. Literture Review [.………………………………………………………………………………………19](#_bookmark2)
   1. Brands Marketing on Social media [……………………………………………………………….19](#_bookmark10)
   2. [Social media and its Consumers.....………………………………………………………………..20](#_bookmark12)
   3. Social Media encourge towars Fashion Brands [………………………………………………….20](#_bookmark15)
   4. [Brand Trust and Loyalty….………………………………………………………………………..21](#_bookmark19)
   5. [Big Data Analytics through Social Media.....……………………………………………………...22](#_bookmark19)
   6. Social Media Marketing Capability [………………………………………………………………24](#_bookmark19)
   7. Risk of Social Media and Big Data ………………………………………………………………[..24](#_bookmark19)
   8. The Intention in Social Media Privacy ………………………………………………………..[......25](#_bookmark19)
   9. Economic Aspect of Social Media Security………………………………………………………. [25](#_bookmark19)
   10. Machine Learning in Social Media Platform ...……………………………………………….[.....26](#_bookmark19)
   11. Supervisory Control and Data Acquistion [….……………………………………………………26](#_bookmark19)
   12. Artificial Neural Network Impact on Social Media[……………………………………………27](#_bookmark19)
   13. Applications of ANN …………………………………………………………………………..[...28](#_bookmark19)
   14. Data Mining in Intrusion Detection ……………………………………………………………[.29](#_bookmark19)
   15. Go Deeper with Convolutions..………………………...…..…………………………………..[...30](#_bookmark19)
   16. CNN Importance in Social Media Field ..………………………………………………………[.31](#_bookmark19)
   17. Supervised Machine Learning Algorithm……….. [……………………………….……………31](#_bookmark19)
   18. Support Vector Machine …..…………………………………………………………………..[...31](#_bookmark19)

CHAPTER 3  [32](#_bookmark2)

1. Project Design and Methodlogy [...……………………………………………………………………32](#_bookmark2)
   1. Information Extraction for Social Media [..………………………………………………………32](#_bookmark23)
      1. Noisy Text Filtering ……………………[…………………………………………………33](#_bookmark23)
      2. Name Entity Extraction ..………………[…………………………………………………33](#_bookmark23)
      3. Name Entity Disambiguatio.……………[…………………………………………………33](#_bookmark23)
   2. Lithium NLP [...……………………………………………………………………………………..34](#_bookmark24)
   3. Dataset…….. [...……………………………………………………………………………………..35](#_bookmark24)
   4. Mathematical Representation (Naïve Bayes) […………………………………………………….35](#_bookmark24)
   5. Name Entity Recognition (NER) [...………………………………………………………………..36](#_bookmark24)
   6. Pattern-Based Method (PBM) […………………………………………………………………….36](#_bookmark24)
   7. Rule-based Method [(RBM)………………………………………………………………………...37](#_bookmark24)
   8. Statistical Learning Method (SLM) [……………………………………………………………....37](#_bookmark24)
   9. Maximum Likelihood Estimation Method (MLEM) [..…………………………………………...37](#_bookmark24)
   10. Hidden Markov Model (HMM) [……...…………………………………………………………...37](#_bookmark24)
   11. Support Vector Machines (SVM) [………………………………………………………………...38](#_bookmark24)
   12. Relation Extraction [……………...………………………………………………………………...38](#_bookmark24)
   13. Supervised Learning Models [……………………………………………………………………..38](#_bookmark24)
   14. Feature-Based Classification Method (FBCM) [..………………………………………………..38](#_bookmark24)
   15. Semi-supervised Learning Methods [……………………………………………………………..39](#_bookmark24)
   16. Multi-Instance Learning (MIL)[……………...………………………………………………….. 39](#_bookmark24)
   17. Bootstrapping Methods (BM) …..[………………………………………………………………..39](#_bookmark24)
   18. Graph Based Method (GBM) [......………………………………………………………………..39](#_bookmark24)
   19. Unsupervised Learning Models […...……………………………………………………………..39](#_bookmark24)
   20. Clustering Methods (CM) .…………………….[..………………………………………………..40](#_bookmark24)
   21. Open Information Extration (Open IE) [………….……………………………………………..40](#_bookmark24)
   22. Wrapper Induction Method (WIM)[……………...…………………………………………….. 40](#_bookmark24)

CHAPTER 4  [41](#_bookmark2)

1. [Results and Discussion………….…………………………………………………………………..…41](#_bookmark30)
   1. Data Preprocessing.………….………………….[..………………………………………………..41](#_bookmark24)
   2. Nature Language Processing [………………...…….……………………………………………..42](#_bookmark24)

CHAPTER 5  [44](#_bookmark2)

1. Conclusion […………...………….…………………………………………………………………..…44](#_bookmark30)

REFERENCES  [45](#_bookmark2)

**List of Figures:**

[Figure 1.1: Graphical Representation of Social Media Usage 08](#_bookmark11)

[Figure 1.2 : Effects on Social Media related to the Advertisement 13](#_bookmark13)

[Figure 1.3: Credibility of Product on Social Media Platform 15](#_bookmark17)

[Figure 1.4: Gantt Chart 17](#_bookmark21)

[Figure 3.1: State-of-the-art vs Novel Proposed Methodology 33](#_bookmark32)

[Figure 3.2: Lithium NLP Pipeline Overview 34](#_bookmark33)

[Figure 3.3: Example working of Lithium NLP 35](#_bookmark34)

[Figure 3.4: Popular Techniques for IE the Datasets used in this research 36](#_bookmark36)

[Figure 4.1: Sample of the Dataset 41](#_bookmark33)

**List of Tables:**

[Table 1.1: Risk Management 18](#_bookmark11)

# **Abstract**

The involvement of social media with every aspect of life i.e., politics, entertainment, education, and much more, are now scenarios that require a better way to search and extract information from social media due to a huge amount of data traffic and content on it. The fashion industry, like most businesses in trendy, has been closely impacted in current years by way of the creation of online social media. Social media has modified how human now not handiest talk however additionally how firms put it up for sale and attain their purchasers. There are 4 crucial social media retailers currently dominating the market: Facebook, Twitter, Pinterest, and Instagram. It might in large part benefit agencies to focus on developing a presence on these websites. Whereas Facebook and Twitter were mainstays for pretty some time now, Pinterest and Instagram allow a great possibility for commercial enterprise increase and innovation. Specifically, Facebook applies this to extract information and posts from the domain of Fashion. The requirement of having a better way to search and extract information from Facebook on Fashion has advantages for content creation and writing, or even just referring to a news article that went viral on Facebook. The purpose of this research extracting fashion-related news from Facebook enabled with machine learning to explore the ways social media influences fashion consumption of the male and female. The primary motivator for the usage of social media is to live in contact with friends and community.

**Keywords:** Machine Learning, Social Media, Facebook, and Information Extraction

**CHAPTER 1**

1. **Introduction**

The history of the Internet is not too old for the world, it's been 30 to 40 decades since the revolutionaries of the Internet captures the world's attention and makes it a global village. Long-distance interaction with friends and families has been a great concern for people for centuries. The roots of social media started flourishing when the first email was sent in 1971. This creates the need for long-distance communications, gradually social media becomes the major source of communication among people living not only in a similar location but also around the world. Social media becomes more open and easy to use due to the easy availability of smartphones, tablets, and portable computer machines including a notebook, laptops. Since the early 90’s the internet has had a ground-breaking impact on society in general and more importantly on businesses. The Internet plays a big, and increasingly important role in global business. Almost 75% of the world has Internet access, while access to the Internet using a mobile phone more than doubled. Virtually all businesses communicate with their suppliers and customers via the Web and email.

The fast development in IT over the most recent twenty years has prompted development in the measure of data accessible on the World Wide Web. Another style for trading and sharing data is web-based media. Online media alludes to the methods for communication among individuals wherein they make, offer, and trade data and thoughts in virtual networks and organizations (like Twitter and Facebook). In today's era, social media users have crossed billions or even trillions in numbers, this means that every single user is creating a bulk amount of data from its account (Badieh Habib Morgan and van Keulen, 2014).

Online media has gotten one of the significant methods for correspondence and substance creation. Thus, modern frameworks that can deal with rich client-created content from the web-based media stage have a few genuine applications. Besides, because of the substance style, size, and heterogeneity of data (e.g., text, emoji’s, hashtags, and so on) accessible via web-based media, novel NLP procedures and frameworks that are planned explicitly for such substance and can conceivably incorporate or take in data from various sources are profoundly valuable and appropriate (Bhargava et al., 2017).

**1.1 Social Media Impact on Society**

It’s an apparent fact that web-based total media policies several aspects of existence as a long way as we might be involved. As expressed in advance, inside the previous decade especially, interpersonal interaction has evolved dramatically. To mention that the enterprise has bounce might be putting it mildly. It has changed the sector we stay in. whilst there are various one of a kind media locations, Facebook, Twitter, and most as of overdue Instagram and Pinterest have taken off and become critical for the ordinary lifestyles for some (Robertson). One should say these 4 locales weigh down the market and typify the thoughts and things one partner with "online media" in dialogue. Figure 1.1 describes the graphical representation of social media usage.

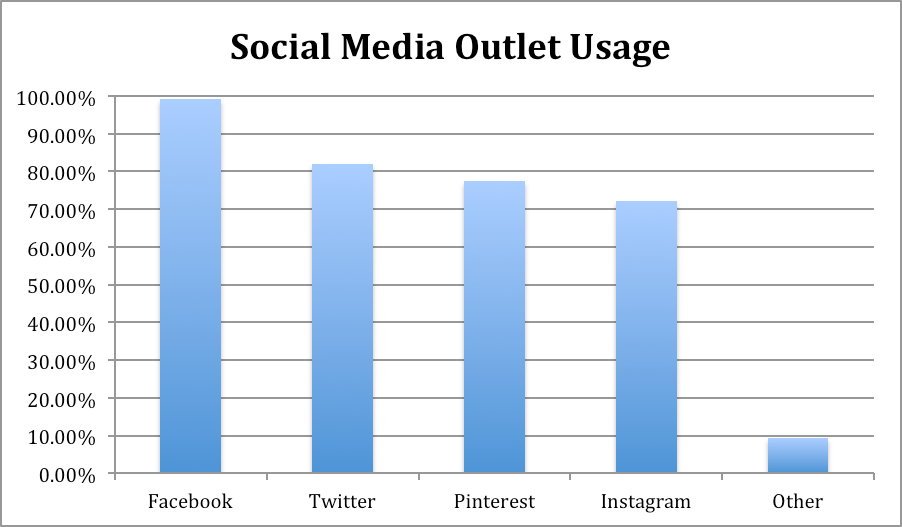


Figure 1.1: Graphical Representation of Social Media Usage

In beneath ten years, Facebook evolved from a "fledging startup to a $114 billion monster" (IBTimes personnel). Due to its preliminary presentation, FB overwhelms the net-primarily based media world today and embodies all that the commercial enterprise entails. It’s far a gap to live in contact with partners, percentage content material, and even gather news and facts. Albeit net-based total media locales existed preceding Facebook, they have been without delay ousted as FB got force. As an instance, in 2006, Myspace changed into the focal point of the online media universe. Besides through 2009 Facebook had extra information open and become developing quickly. Perhaps people had been so interested in Facebook in its preliminary years because of the freshness and energy encompassing the advancement. notwithstanding, in contrast to big numbers of its adversaries, Facebook would not look like going anywhere, and what has saved fanatics trustworthy for such endless years is something notably greater mind-boggling than some other layout. As indicated by using ongoing exploration "two new examinations have tracked down that the informal network fulfills fundamental human cravings, each frame and soul”.

The organization may additionally owe a part of its prosperity to the best, trust wonderful country it conjures in its clients. Exploration discoveries additionally uncovered that crucial social requirements may be answerable for people over and over travelling the web page. The number one need is the focal craving to have a place. Human beings are wired to companion with unique people, too to sense stated by means of them. Facebook empowers those associations in tiers. (Sheldon et al., 2008). Feeling disengaged in the beginning persuades human beings to make use of the web page (basically as a method for handling strain), and accordingly, the more habitually any person utilizes the site, the extra associations they collect. "The second one friendly thought fb reacts to be self-show, in which client’s style perfect of themselves through their profiles". That is shown in, as an example, the amount of 'companions' of their organization and pix. Therefore, FB personas which might be brought on-line might be an all the greater socially engaging 'self' this is attempted to surely, however one that isn't always yet stated (Zhao and Zhao 2008)".

Through this exploration, it is not tough to get take care of the truth of why Facebook has gotten the main staple of everyday existence. Individuals respect feeling in charge of their universes and FB is unequivocally the spot to do as such. In May 2012, Facebook's First sale of inventory turned tremendously disappointing. In any case, all in all, this meant "flop" has not appeared to avoid the website's online prosperity. In light of customers' almost strict propensity, the web page is as but prospering and it appears to be a simple function that an ever-increasing number of organizations could ultimately get on board with and use the web page for their potential advantage. Whilst FB at present administers the sphere, Twitter is some other outlet that immediately gathered impulse. He was a software program product designer composing dispatch programming but "were given captivated with the aid of the way that everything of the clients of the product cabbies, limo drivers, messengers were without a doubt detailing what they had been doing, exposing down essential realities simply, therefore the idea for Twitter become imagined." (Carlson). Through an extended arrangement of activities, Dorsey ultimately was given related to a tech-new business and made his idea of a standing broadcasting framework a truth. Referred to then as "Twitter", it became a framework wherein one could send an e-book to a solitary wide variety and it's communicated to a large range of partners. Dorsey stated, "We ran over the phrase 'twitter', and it turned into truly splendid. The definition becomes a brief eruption of insignificant facts', and 'peeps from birds.' And that is utilizing and huge what the object becomes" (Carlson).

A normal grumbling in regards to FB is regularly that it's miles just "to an intense." some organizations do not want to be besieged with photographs and indulgent notices and Twitter to a wonderful quantity killed this option. Considering that just a hundred and forty characters are accredited per tweet, account holders can appear over the web page efficiently through a laptop, tablet, or PDA and burn through extraordinary measures of statistics in a reasonably brief time body duration. Some other advantage Twitter offers is the way that one doesn't need to "tweet" or make a page-to technique - although it is hard to observe FB without setting up your very personal page in advance. Even as many can also see Twitter as only an aspect mission of FB, its various traits are probably in general chargeable for its developing success. "Seat Exploration center's figures propose it is going to be difficult for Twitter to arrive at Facebook billion-component level. Nonetheless, over the maximum recent 9 months of 2012, Twitter customers extended through over 40%" (Berman). Twitter's vast benefits lie in classified ads in which organizations are normally paying to be talked or "tweeted" about. Berman explains and states, "it hanging because Twitter has been in the rewarding game for simply three years, basically selling " backed tweets" to publicists whose come-ons spring up within the message floods of the service's 200 million or greater dynamic customers" and incomes are pushing the billion-greenback mark. (Berman) As expressed in advance, by way of and large, the distinction of businesses using online media is that there is reasonably 0 overhead fee. on no account like procuring area in a paper, shopping for time on television for a commercial enterprise, or renting an assertion, is admittance to net-based totally media free. Berman said all that needed to be stated while he expressed, "That attraction is the well worth of free work. The agency doesn’t want to pay a penny for the 400 million messages sent each day. Likewise with FB, the clients energetically, ravenously, accomplish the work complimentary" (Berman).

While FB and Twitter currently look like almost essential wares, Instagram is the furthest down the line fever to revel in the spotlight. As of late procured via Facebook, "Instagram is an online photo sharing and informal verbal exchange management that empowers its customers to take snap shots, observe computerized channels to them, and provide them on a collection of interpersonal interaction administrations, for instance, media destinations along with Facebook or Twitter" (what is Instagram?). To a few, it has a tendency to be regarded as the Polaroid of our age; photos are even explicitly designed for iPhone similarity. even as Twitter took the attention a part of Facebook and exploited it, Instagram has thoroughly done likewise with the photo-sharing section and declared on February 26, 2013, that it arrived at 100 million month to month clients, "this means that a hundred million people are utilizing Instagram at any rate once according to month and plenty of are probably utilizing it all the greater regularly" (Tinari, 2016). Furthermore, more or less forty million photos are published every day, along with web page traffic which means 8500 "likes" every 2d, and one thousand feedback each 2nd (Instagram Press).

As per Douglas (2016) a supporter of CNN, "Instagram has ascended to the extent of a Twitter all things considered. It has generated every other visible language, another behavior of sharing, and an overflowing of creativeness as challenges, network workmanship displays, and individual articulation. Instagram got so well-known so brief that FB paid heed and bought the business enterprise for one thousand million dollars to assist its advanced cell cellphone presence presently before its initial public offering".

Insights like those are what exhibit Instagram is the upgraded "it" issue and groups are rapidly paying heed. Instagram offers a degree to businesses to present clients an within the heritage look and restrictive admittance to their organization along these traces causing the consumer to experience extra associated and preferably bound to hold helping the brand. Inside the retail global, numerous agencies are using Instagram as a form of e-inventory, anyway, this will now not be the best method. "Close to Instagram, devotees want to look in the historic photos from their primary organizations that they could not more often than not have the choice to via customary media. As an instance, (Burberry, 2015) in comparison to its top-of-the-line contenders, makes use of its file to post cool photographs of London, where the business enterprise is based totally, along with within the back of the stage pictures from picture shoots. It posts an intermittent product image, yet doesn't flood its document with laborious images of garments" (Austin, 2017). The motive of utilizing Instagram as a manner to raise commercial enterprise is to provide customers something they broadly speaking could not revel in from on foot into a store or on any occasion, cruising a site. The objective is to purpose supporters to sense as even though they're essential for the business enterprise and are "tuned in" with any large happenings previous to something goes fashionable.

**1.2 Fashion Industry Target through Social Media**

As indicated through the Reference eBook Britannica, "the design commercial enterprise is a multibillion-dollar worldwide assignment committed to the problem of creating and selling garments. Some eyewitnesses apprehend the design enterprise (which makes 'high fashion') and the attire commercial enterprise (which makes common garments or 'mass fashion'), but by way of the 1970s, the limits among them had obscured. The layout is exceptionally characterized essentially because the style or forms of getting dressed and frill worn at some random time by way of gatherings of people." (Steel and most important). It’s miles apparent these days that the style commercial enterprise assumes an enormous component within the public eye. People use style as an average of articulation and a wellspring of self-character. There may additionally seem like a generous hole between the steeply-priced, restrictive high fashion of New York and Paris and the mass-created easygoing wear offered in shops of the suburbs even though, all degrees of the style enterprise fuse similar thoughts of "plan, fabricating, flow, showcasing, retailing, publicizing, and advancement of an extensive range of apparel (guys, ladies, and children's) from the thinnest and steeply-priced excessive fashion (in a real experience, 'high stitching') and fashioner designs to normal garb—from couture ball outfits to Succulent Couture-logo going for walks pants" (steel and predominant).

to show the weight the layout commercial enterprise holds inside the gift area, in 2010, the arena attire and fabric enterprise came to almost $2,560 trillion and the girls' wear vicinity on my own is relied upon to skip $621 billion utilizing 2014 (Breyer, 2017). In completely three hundred and sixty-five days, the USA burned via $250 Billion on style usage (design enterprise Measurements). These sizable dollar sums explicit how massive of a task the fashion business performs into the usage examples of people all over the planet.

Inside the case of reasoning sensibly, it simply bodes nicely to expect the style commercial enterprise will constantly exist. There are more than 7 billion individuals in the world and clothes if taken into consideration cutting-edge will continually be a need, thusly a few interests will consistently exist. Many might also contend that layout is an extravagance and pundits cannot help considering why it is even huge, being as most of the sector cannot manage the price of bona fide couture. Nevertheless, in a brand new meeting on the Colbert document, Anna Wintour, manager in the head of the vogue magazine, protected the enterprise. She stated, "Further than amazing many people who appreciate a Picasso may not ever have the option to get one, maximum of people who pine for a splendid Chanel fit won't ever experience that fleece against their skin. anyways, a good deal as you can anyways like Picasso's pieces and his important commitments to the workmanship global as a top supporter of the cubist improvement, you may anyhow take a gander at early Chanel plans and perceive how subjects of women\'s strengthening and motion were confirmed inside the plans" (LaGrave, 2018).

Wintour is contending that an extra profound appreciation for style includes a comprehension of its patterns, impacts, messages, and history-comparable as compelling artwork and maybe this more noteworthy importance is the thing that continues purchasers spellbound. Likewise, just like expressive arts, it cannot be rejected that high layout is an extravagance. For the reason that the widespread majority cannot stand to revel in, the longing to stumble upon the "appearance" of high layout is moreover heightened. Due to shops like H&M, Everlastingly 21, and even objective, even the maximum cost cognizant client can situate themselves as conscious innovators. In this manner, the intellectual elements of the enterprise no longer solely can legitimize the first elegance, the pinnacle of the line customers, yet similarly the deal trackers. Assorted to before, today's creator’s work in a global commercial middle, constantly extended utilizing television, film, and mainstream society, however extensively in addition accelerated by using the web and on-line media. Now not, at this point restricted to the privileged, pioneers come from one of a kind social corporations, and maximum patterns have quick instructions. Figure 1.2 describe the social media impact and also used for advertisement purpose.

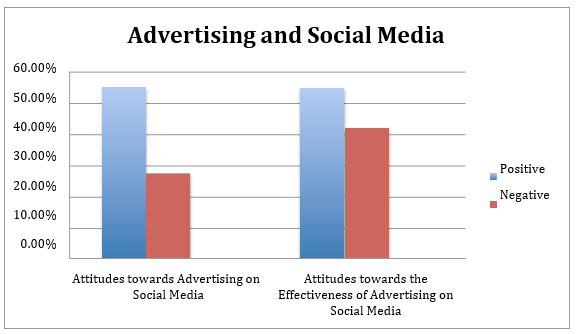


Figure 1.2: Effects on Social Media related to the Advertisement

Jennifer (2019) observed that the cerebrum cherishes new things, in particular, drifts due to the fact they may be rapid glimmers of interest. Comparable as repainting a room, or shopping for new fixtures, the acquisition of every other, famous thing considers reexamination, and the purchase of a wearable elegant thing takes under consideration self-reevaluation (Baumgartner, 2018). Likewise, as expressed previously, fashion is continually being suffering from its preferred weather be chronicled occasions, mechanical headways, or converting social jobs, and sporting this layout is a strategy for articulation. "All get dressed selections are inside propelled and may be investigated to discover the internal identification. Patterns pursuing is often spurred from a craving to healthy in, sense contemporary, and veils uncertainties. These continually changing design directs never think about a sizeable id of your look, your inclination, and at closing the message you want to develop on this planet". Typically, the manner wherein one decides to get dressed is an impact of their maximum genuine self.

Comparable to fashion, web-primarily based media is any other outlet to articulate one's thoughts. Be that as it could, this method for articulation takes place inside the net. Numerous fundamental strings join the 2 together. For example, the intellectual attitude to introduce oneself with a selected intention in thoughts exists inside the two retailers. individuals get dressed in a selected technique to introduce themselves to the world around there, almost equivalent to the manner wherein customers might also mastermind their Fb page, Twitter record, or Instagram. Furthermore, the craving to healthy in and be incorporated gives itself through design and internet-based total media. Customers of fashion need to be a chunk of the maximum latest sample and get awards from friends, similar as customers of internet-primarily based media want their "partners" or "adherents" to like and comment on their posts. At remaining, speaking to most of the people is also something the 2 retailers preserve in like manner. These days, there are extra than 1 billion clients of web-based media and minority holes are essentially non-existent. Shockingly, no longer utilizing online media now frequently compares to a beneficiary of elitism, no longer the failure to get to it, as before."Stylish human beings find out it too standard and others find out their protection arrangements irksome. All in all, not utilizing net-based media is probably a result of more education, no longer an absence of getting entry to" (Ferenstein, 2018). Figure 1.3 showing the credibility of the product in different forms or levels.

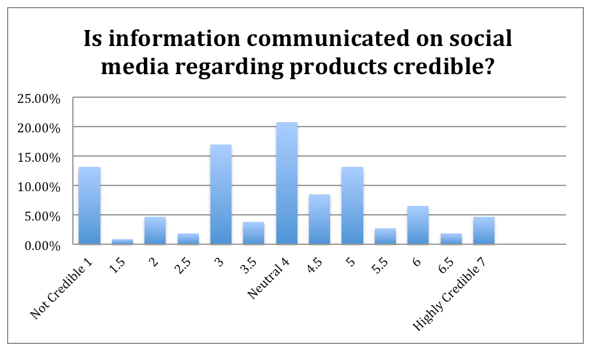


Figure 1.3: Credibility of Product on Social Media Platform

Concerning layout utilization and online media in Oxford, Mississippi, it seems as even though Oxford may perhaps be an anomaly. Demographically speak me, Oxford is a touch, southern, faculty metropolis; thusly it doesn't nicely deal with the more prominent US so any ends that came to may not be suitable to the complete layout retail variety. Of the surmised 16,000 understudies, 55% are female and nearly 37% of the understudy frame is out-of-express. A big level of the college is Greek associated, and as expressed inside the presentation, understudies are very look arranged; numerous with plain tiers of more money. It seems to be like this profoundly public manner of lifestyles, blended with youngsters based through on-line media with the capability to keep, could sincerely impact close by retailer's selling techniques. Accentuating a presence thru on line media could be worthwhile.

**1.3 Background of the study**

At some stage in the most current decade, the computerized advancements have notably modified the approach of consumer's correspondence with one another, the technique of finding and buying and selling statistics approximately items, and the approach of their purchasing and devouring. Net is currently a substantial wellspring of statistics that assists customers with selecting all the extra efficiently during their buying (Valerio, 2014). the new media channels, as an example, Facebook, YouTube, Google, and Twitter set the buyers in a more dynamic task as market players and empowered them to attain all various wherever and every time (Thurau et. al, 2010).

The customers of the casual agency may additionally collaborate with the organization from a couple of factors of view: via imparting insights, making sizable substance and information, assisting showcasing efforts, and so forth, however, they can likewise talk with each other. Those institutions are important from the standpoint of agencies because such dynamic clients construct the emblem with the aid of expanding mindfulness, inclusion, and dedication, and hence they invigorate buys. Hence, organizations need to always foster better procedures to companion with their customers, bring together strong connections, and increment social dedication to drive improvement (Mohr, 2013). The coming of net-based media because the harbinger of records to all people all for the duration of the planet efficiently and unbiased old enough has been stated unknowingly by way of all on-line media clients and they put it to use. Facebook in such a manner has been flourishing because of it being an early internet-based media with a ton of records and following. With people posting the maximum recent information on all styles and occasions when they find out about it, news about the maximum latest patterns in design will likewise be on FB.

**1.4 Objectives**

Using Machine Learning, this project aims to automate the process of the classification of Facebook posts into Fashion and non-Fashion posts. This would be achieved through the following steps:

* **Dataset selection:** Selecting a dataset that contains posts from facebook related to the topic of Fashion and is properly labelled to satisfy the requisites of this research.
* **Data Preprocessing:** The dataset selected for this research contains textual data. Therefore, appropriate data preprocessing is required to make the data usable for training classification algorithms. Natural Language Processing techniques namely the Count Vectorizer and the TF-IDF transformer will be applied to extract the hidden features from the textual data for training the algorithms. The dataset will also be split into training and test datasets for cross validation.
* **Machine Learning Algorithm:** Upon analyzing the problem statement, the Machine Learning paradigm for this research will be binary classification since the aim is to identify and classify whether a Facebook post is Fashion related or not. Three algorithms for this purpose are selected, the Multinomial Naïve Bayes algorithm, the Support Vector Machine Classifier and the Multilayer Perceptron.
* **Performance Metrics:** Once the Machine Learning algorithms are trained, their performance will be measured using performance metrics meant for classification algorithms. These metrics will provide graphical and empirical data on the performance of each algorithm which will then help identify the best algorithm for the classification of Facebook posts into Fashion and Non-fashion posts.

**1.5 Research Question**

The following research question will give the project direction:

* How can Machine Learning make surfing on Facebook for a particular topic/niche easier for the users?

**1.6 Ethical Considerations**

Using data that belongs to a random user on Facebook without permission is as unethical as it appears to be. However, datasets with this information that are publicly available require a different approach. Ethics is a sensitive issue and developers working with anything related to Facebook, whether it be their APIs or their Frameworks or their private data, they have clearly devised Developer Policies for all developers that wish to use resources from Facebook for their needs and requisites. The Developer Policies related to this research are as follows:

* Product Quality
  + Anything related to the developer’s needs must be accessed using the Developer’s own Facebook Account.
  + If an App is being built, the stability and ease of navigation must be guaranteed.
  + If the app being developed requires the use of a bot, then the bot’s description and categorization must be kept updated.
  + Causing any kind of confusion, deception, fraud, misleading and spam if forbidden.
  + Facebook expects negative feedback on the app if it is being developed, however Facebook employs a threshold for this and the negative feedbacks must not cross that threshold.
  + Apps with minimal utility that provide predictions, assessments, or similar outputs to the user, may not be allowed on Platform. For example, Apps that provide (or claim to provide) users with assessments of personality, personal attributes, character traits, behavioral tendencies, or whose core functionality otherwise involves making predictions about who the user is, may not be allowed.
* Give People Control
  + Getting the permission of the user whose data is being used is mandatory.
  + Don’t prefill any content in captions, comments, messages or the user message parameter of posts unless (a) it is a single hashtag in a post shared through our Share Dialog (but not via our APIs), (b) it was created by the person using your App, or (c) it was created by a business whose employees use your App to administer the business’s presence on Facebook.
* Encourage Proper Usage
  + Be honest about your relationship with Facebook when talking to the press or users. Comply with our Developer PR Guidelines and get approval from us before issuing any formal press release or blog post mentioning Facebook.

**1.7 Project Timeline**

The Gantt chart is identified as a strong tool for time management. The Gantt chart designed for this project is laid out below in Figure 1.4. Tasks are laid out in chronological order on the left-hand side. The timeframe for their completion is found along the X-Axis. By sticking to this schedule, the project will be delivered promptly to a high standard.



Figure 1.4: Gantt chart

There may be unexpected events that cause difficulty in maintaining the project timeline. These are risks. Contingencies for some expected risks are discussed below in Table 1.1.

|  |  |  |  |
| --- | --- | --- | --- |
| **Risk** | **Risk Impact (1-5)** | **Risk Likelihood (1-5)** | **Contingency** |
| Data Mismanagement | 5 | 1 | Each iteration of the dataset will be synced with Version Control. |
| Personal Time Mismanagement | 3 | 3 | Regular Feedback from Supervisor for Continuous development. |
| Hardware Failure | 1 | 1 | Extra Supply of RAM and SSD. |

Table 1.1: Risk Management

**CHAPTER 2**

1. **Literature Review**

A strong Literature Review (LR) provides validity to the integrity of the research performed. The studying of existing work done by researchers in the same domain refers to as LR. The following papers have been selected and summarized for this research’s LR.

**2.1 Brands Marketing on Social media**

Kim and Ko (2012) noted by way of Richter and Koch (2007) characterize internet-based totally media as online applications, tiers, and media which work with connections, joint efforts, and the sharing of the substance. Online media is being utilized by almost all agencies paying little heed to their duration or their type of organization to publicize and enhance themselves. The quantity of agencies with a report through internet-primarily based media keeps on developing (kumar and lakshmi, 2012). Several companies start to make use of net-based media in marketing, public correspondence, or a comparative department, in which it is important to assure to have a proper away connection amongst employer and clients. Clients can entice companies for example by way of posting remarks on agencies' profiles. Besides, its miles a method to construct mindfulness among customers who are remote of standard media. To put it clearly, numerous businesses are seeking out customer willpower, and that they see internet-based totally media because of the superb approach to do it (Evans and Mckee, 2010).

Using online media became at the start attempting for layout brands. They had been involved in how their photo might be seen given that being handy thru internet-based total media regarded as masses for them for the purpose that they felt further uncovered (Morriseey, 2010). Not with standing, today's publicizing and showcasing device may be negative without utilizing internet-based media. on the point even as fashion indicates have been near occasions only for customers and press without famous humans inside the first line, isn't always any greater. The assortments are currently introduced now not absolutely to the picked few, but all clients all at some point of the planet because of the cutting facet innovation (Wylie, 2012). Furthermore, design sweethearts, who are starving for the maximum recent styles, do now not want to rely on the ultimate version of the style, because of the reality that online media has colossally modified the condition of favor commercial enterprise. Supporters of FB, Twitter, Instagram, or Pinterest money owed are overflowed with live streams, tweets, posts, and pins and they are informed about the maximum current styles proper away (Wwd, 2013). The previous single path correspondence thru standard media turned into supplanted thru imaginative smart-way correspondence and along those lines fashion groups that don't display up via internet-primarily based media, absolutely lose their top hand (Costa, 2013).

Brands and customers started to speak with one another without the restriction of time or spot. The intuitive -manner of direct correspondence empowered manufacturers and customers to participate in the improvement of recent devices, administrations, plans of motion, and trends. The principal advantage of net-based media showcasing is the hazard to decrease misjudging and bias toward brands and increment logo esteem thru constructing up a stage, wherein individuals can change their minds and information amongst themselves on the net (Kim and Ko, 2012).

## **Social Media and its Consumers**

finally, of the maximum extremely-current decade the automated developments drastically have modified the way wherein buyers communicate with one another, the approach of locating and buying and promoting records about devices, and the approach of their buying and devouring. New media channels, for instance, FB, Youtube, Google, and Twitter positioned the customers in a greater dynamic task as marketplace game enthusiasts, and empowered clients to gain (and be reached through) anyone anywhere and whenever (Thurau et al., 2010). Statistics innovation is appealing to clients and their interest is converting from latent beneficiaries of information to dynamic turbines of statistics. Sports activities as quickly as restrained using companies begin to be performed increasingly more with the aid of customers (Heinonen, 2011). As a result, the agencies fostered a need to understand this adjustment of request to have the choice to benefit from the use of internet-based media.

## **Social Media encourage towards Fashion Brands**

Parvinen, and Malmivaara (2013) inspect thought for utilizing enterprise community area web page from a changed attitude. The exam inspects significant components of the muse for making use of brand neighborhood place pages: indulgent and utilitarian. . The clients with a utilitarian suggestion need to arrive at a selected intention via the nearby location, like coming across supportive records previous to deciding on a buying desire, even though gluttonous inspirations are diagnosed with seeking out no specific cause and entertainment. The utilitarian idea is associated with accepting treasured information about the element of hobby and helping with deciding on preference recognized with utilization. Alternatively, the gluttonous idea is hooked up with searching through a way to invest loose electricity.

In mild of Jakste and kaite (2012) customers begin to attract with the producers the maximum due to acquired social advantages. Customers are roused to have interaction with online networks with the aid of their longing to find out approximately an emblem or a selected object, just as to obtain help, to be a bit of the nearby region, or to get an opportunity to communicate their feelings and present themselves. Jakste and kaite (2012) likewise examine that customers can attract with a logo through their inward idea to face out sufficient to be located and connect to nearby location humans. Thurau et al. (2004) sum up 8 explicit elements that rouse customers to draw in with online networks: (1) ventilating negative sentiments, (2) worry for one-of-a-kind customers, (3) self-improve, (4) exhortation chasing, (five) social blessings, (6) monetary blessings (rate reserve budget), (7) level help and (eight) helping the organization. The exam advises that from the consumer's factor concerning view, the possibility of social blessings has a pleasant impact, all matters considered.

The examination directed by using Baird and Parasnis (2011) reveals that there may be a hole among what businesses assume customers are eager on and what customers want from their internet-based media connections as a popular rule. Consumers invest their power, give individual information approximately themselves and they can even help the brand. Along these traces, they're looking ahead to some aspect substantial as an alternate-off for that. The exam uncovers that most customers talk with groups thru net-based total media when they need to get a markdown or to buy objects. Expertise audits and object scores changed into referenced at the third spot. To build up well-known and selective records changed into additionally a concept referenced amongst exceptional functions for connection with an emblem via online media. The 6th spot changed into worried with the aid of using the craving to discover new objects.

## **Brand Trust and Loyalty**

Sashi (2012) suggests that the intuitiveness of online media works with the constructing up of withstanding connections amongst customers and brands. Moreover, as in line with him, the intelligence includes the threat to maintaining up the dialogue the various emblem and consumers, who become related to the advent of their substance. Gummerus et al. (2012) affirm that emblem networks carry freedom to companies and clients to draw in with each other. at the same time as groups focus on drawing in with steadfast customers, influencing humans' discernments about the logo, spreading mindfulness about the emblem, and gaining from and approximately clients, clients get blessings via the collection of practices that they do on the internet. Brodie et al. (2013) suggest a few effects that can upward push out of patron determination like agree with, achievement, responsibility, enthusiastic affiliation, and unwaveringness. The studies stress that the online logo nearby location is for the maximum issue diagnosed with steadfastness, responsibility, and strengthening. Gummerus et al. (2012) see devotion as a focal justification for brand nearby place collaboration. He offers that clients might be a part of the brand nearby region when you consider that they just like the emblem and thusly they experience devotion to it. Because of consumer determination dependability can come to be extra robust. It modified into established that insofar as fans and adherents are extra sincere, they commonly purchase effects of the logo that they examine, and besides they'll prescribe it to exclusive humans.

Laroche et al., (2013) stated concurring Chaudhuri and Holbrook (2001) delimit brand consider as "the fervor of the everyday client to have confidence inside the capacity of the logo to do its expressed capability." trust within the brand and in a while faithfulness to it are extensive for purchase expectation. Kim and Ko (2012) symbolize purchase expectation as an aggregate of consumers' benefit and the possibility of purchasing an item. The buy purpose is firmly recognized with the customer's disposition and tendencies towards a logo. Muk et al., (2014) cited by using Zeithaml (1988) upload that buys expectation is the chance of customers' readiness to acquire an object. Corcoran (2014) specifies that there's a connection amongst income and web-primarily based media just while the company has drawn in the gathering of fans, who are always conversing with one another. In the end, at the off chance that the relationship develops into steadfastness and devotion to the brand and it can affect income and offers.

Even though promoting and buying items is the center of each commercial business enterprise. Kim and Ko (2012) name attention to that the in reality worth a business organization reaps from customers isn't genuinely restrained using the pay from every trade besides that the whole advantage contains the customer's movement he can offer over the time of his dating with the company. In this manner, clients may be assessed as immaterial assets of the commercial enterprise employer, which ought to acquire equal consideration as other financial property.

## **Big Data Analytics through Social Media**

Thakare et al., (2010) reported that to gain suitable knowledge from logically related datasets, scientists have discovered a new domain that is known as data mining i.e. extraction of unknown and useful patterns, information from data. But data mining techniques cannot be helpful always to mine data sets. Big Data is a term that is used to identify larger datasets i.e. larger in size and complexity. These datasets are difficult to mine because they can't get stored, manage, and analyze easily with data mining software or previous tools and technologies. The reason is Big Data sets are heterogeneous i.e. structured or unstructured and extracting useful information from these larger sets or data streams was not possible due to their variety, velocity, and volume. Analyzing abundant data comes under Big Data analytics, which is the procedure of smearing progressive analytics and visualization tools and techniques to determine unknown patterns, correlations that help in authentic decision-making sense. Big Data analytics is a multi-stage process that involves data attainment and recording, information abstraction, cleaning, integration, aggregation and representation, processing query, modeling data, and explanation. Every stage of analysis comprises various other processing tasks so there come many issues and challenges in mining larger datasets. These issues and challenges majorly include (Liu, 2012).

1. Heterogeneity and Incompleteness: As Big Data normally deals with structured and unstructured data, this creates a great challenge to mine such huge data. In complicated heterogeneous data (data in multiple formats), data sets can have multiple business rules and patterns implications by concerned organizations. It may have no particular format like it may include formats like emails, videos, audio, graphics, x-rays, textual, such formats cannot be stored in structured databases. So there arise issues to transform such datasets into a suitable format. Whereas incomplete datasets or missing values are also a great challenge in analytics.
2. Scale and Complexity: Scale represents the growth rate and volume of datasets. Traditional analytical and mining techniques couldn't afford the scale and complexity of such complex datasets.
3. Timeliness: Time is a greater challenge in analyzing larger datasets, this is because a requirement for analysis expects the results within shorter intervals but the fact is, it is a time taking task.
4. Security and Privacy: Normally organizations don't want to share their data with other companies. If the company is analyzing its own data sets from its employees then it might not be an issue but if a third party is involved in the analysis phase then there arises huge security and privacy concerns.

The business information system has been consumed with the data caused by social networking that raised the requirement to absorb that into their business tasks for superior decision making in the era of big data. Even so, such new streaming of data, generous, unshaped, and variation, bought manages the data warehousing of the structure, and merging of gadgets to their joints have inspired us to perform such research work (Jenhan et al., 2019).

We have advanced and narrated a novel proposed system for the big data implementations that can issue recommendations on the behalf of the reaction of the users that creates multimedia satisfaction among more than one user of social networks. The system depends on the 'User centered' attitude (Amato et al., 2019).

Inspecting and removing the understanding of the user-generated facts has become a matter of attraction between businesses and the investigation team because such facts consist of precious information that could involve viewpoints of consumers, ranking, and guidance of outcomes and solutions (Chang et al., 2019).

Investigation of big data is anticipating research instead of the customary illustrative study of data. Hadoop is been used almost universally for an investigation of big data in social networks like Google, Gmail, Amazon, Instagram. The distributed file process is been used by Hadoop for the savings of large proportions of unshaped, present data, and flood at excessive velocity Suganya and Selvamuthukumaran (2018).

## **Social Media Marketing Capability**

In this examination, Basha et al., (2019) web-based media advertising abilities are conceptualized as the capability of firms to serve client needs of obtaining and sharing legitimate, solid, and ideal data through viable and brief online exercises. In a quickly developing and data-driven commercial center, client hazards are constantly intermixed with web advancements and data sharing. It is especially obvious in the list of online protection, which demonstrates that media and public insight hazard to the corporate data is the most conspicuous impact of hazard to technology. The informal community hypothesis improves our comprehension of how irregular individuals are interconnected and interrelated. It places that connections instead of people are significant in an organization. Predictable with dynamic capacities hypothesis and informal organization hypothesis, this examination conceptualizes a powerful promoting ability, specifically, web-based media advertising capacity, as a capacity of the firm to reconfigure and increase its innovative, educational, and social assets to produce positive partner-based affiliations.

## **Risk of Social Media and Big Data**

Subroto and Apriyana (2019) described an algorithm model which is used to utilize big data analytics and statistical machine learning to predict the risk of cyber security around social media. Data were collected from early 1999 to 2017 and 25599 cases of cybercrime from Twitter. Prediction of cyber threats through analyzing the software vulnerabilities based on the conversation of social media. In this research mention, some happened cybercrime in 2014 and these crimes are, JPMorgan Chase lost the 250 million us dollars approximately, Sony Pictures Entertainment company lose the 100 million US dollars, Target loss 1.2 million US dollars and the company name Home Department was the loss the 90 million US dollars.

## **The intention in Social Media Privacy**

Dlamini et al. (2019) described that information security is not only a concern to the era of mainframe computers it’s now up to the current state of the complex internet through this research describe the evaluation of information security breaches, came from, where it is today and where it is moving. Now a day’s new risks occurred because of the enhancement and inventions of the technologies. In this survey, the paper shows that the information technologies are winded and need to make more security strategies for cybercrimes through national and international society and also present in this survey paper that security challenges are larger extent related to the security of human and organizational aspect so, all these circumstances the new researcher must introduce new technologies for controlling the cybercrimes or minimize the previous gaps.

## **Economic Aspect of Social Media Security**

Gordon et al. (2016) introduced the economical aspect of security cost at information security. Authors focusing on the three basic questions first one is what should be the cost of information security breaches, second is what will be the budget of activities related to cyber breaches and the last third question is how an organization spent the money on cyber security and how it is important for their organization or information technology and Detailed on the Workshop on Economics and Information Security (WEIS).

Cazacu et al. (2019) described the unpleasant economic impact due to the lack of security in the architecture of IT gadgets. Need to control all these cyber problems with fast-growing methods and provide a secure atmosphere to the stack holders of the IT industry and setting up Security Operation Centers (SOCs). Privacy Right Clearinghouse (PRC) mentions that in 2018 there are over 8000 data breaches reported since 2005 and over 10 billion records affected due to these breaches and in 2017 the department of IBM security and ponemon Institute present a study that averaged 3.6 million US dollars exceed due to data breaches.

The development of e-learning tools enhances the security for controlling the cybercrimes and for this purpose applied the different framework of the activity theory based on machine learning techniques and algorithms for identifying the risks with the help of different members of SOCs and present their opinions to diminish these risks.

Lawrence et al. (2014) explain the importance of costs of cyber security breaches on the business and the consumers. The primary object of this research to enhance the Gordon-Loeb model to find out, how the optimal investment is suitable for business on information security. The entire system and whole business-critical infrastructure may become to an end due to cyber breaches.

## **Machine learning involvement on Social Media Platform**

Chan and Lippmann (2016) present the machine learning algorithm to analyze social media new because there are a lot of use computer and the internet in daily life and attract people to attempt undesired activities through the computer. Attackers can attack business, the internet of things and personal data they can access the private data and may shut down the whole system. Therefore, now a day’s information Security is not a rare topic for discussing. Traditional software of cyber security needs a lot of manual effort for detecting the threats, dig up the characteristics from the risk, and then encrypt through the security software for detecting the threat or virus. This manual and the time-consuming process will be handled efficiently through the use of machine learning techniques or algorithms. Many researchers invest and introduce many techniques in machine learning for controlling cybercrime more accurately and reliably.

Noor et al. (2019) described the future of social media recent attacks, and applications of the new technologies and challenges with the prime help of Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) in the field of networking or IT. IoT, AI will boost the CA and server fewer applications will be increased under the protection of data which is saver on personal electronics gadgets related to IT Technology. Unfortunately, there is seems to be no end to hackers who want to access personal and unauthorized information and the organizations and utilities will be the victim of cyber-attacks so, must be governed and financial systems take a hard decision to control these cybercrimes.

## **Supervisory Control and Data Acquisition**

Cherdantseva et al. (2016) This survey paper discusses the security of Supervisory Control and Data Acquisition (SCADA) between ten years of 2004 to 2014 and describes the valuation of risk in cyber SCADA. SCADA is originally a system that is the type of Industrial Control System (ICS). SCADA system consists of software and hardware components and network or networks connecting.

ICS has the responsibility to control the process in industrial areas and the sectors related to the Critical National Infrastructure (CNI). SCDA is used to monitor and control the assets over large geographical areas and manage specific equipment of IT or Electronic which is connected through network/internet such as the terminal unit of Master and Remote (MTU & RTU).

The software which is used in SCADA systems is a real-time database, multi-tasking these functions provides synoptic diagram, text, multi-screen, trend analysis, archiving, report generation, logging many more. There are 10 steps mention in the survey paper about the methodology of SCADA these steps are, i) establish system configuration, (ii) identify quantity model, (iii) identify the security requirement of chief phases, (iv) identify system susceptibilities, (v) categorize the susceptibilities on each device, (vi) estimate the time, (vii) produce graph and attack trails, (viii) evaluation leading attack path, (ix) for enhancing and baseline use 3 and 8 steps, (x) compare the result of system and risk reduction. Definition of Risk According to SCADA “risk is a term of the possibility of an assumed threat source misusing a potential vulnerability and the effect successfully exploitation of the vulnerability”. Modern days SCADA has much ability, sophisticated, modernization, and real-time operation, technological advanced, multi-component architecture, and distribution mechanism for cyber threats.

## **Artificial Neural Network Impacts on Social Media**

Abiodun et al., (2018) described that ANN is another model of machine learning (ML) the development of Artificial Neural Network is inspired by the biological neuron system to simulate the structure and base of functionally like a human brain. As per the name of Artificial Neural Network is a combination of three different words first is Artificial which is defined as it a presenting of a real objector some time called human-made and its function is close to that original object. Second is Neural, neural is an adjective of neurons this was originally taken from the brain of human in our brain there are billions of cells which called neuron and fundamentally work like biological neurons many researcher and developer are used many other alternative words like connection base network, parallel distributed processing network, etc.

Artificial Neural Network is the processing based on an algorithm that can be build complex patterns to predict the problems or provide the solution of that problem. The similarity between Artificial Neural Network and Biological Brain or Neuron System, to know about the functionality of ANN must be needed to know how BNS work because the idea of ANN techniques is originated from Biological Brain/Neuron. Many problems which we solved on ANN but Also can have solved previous or alternative methods so why we move on ANN, basically ANN is very suitable and efficient for those problem solvers who want to get significant advantages such as cost, ease of debugging/maintenance, accuracy, time and many more. There is another advantage of ANN to solve the problem by using the lookup table approach. Fundamentally lookup table is used to store all the information for gaining the appropriate result and reference of upcoming events and through lookup table approach to generalize the data and in this generalization, ANN will be trained to provide an appropriate solution to the required problem and ANN train through many inputs according to the problem which we are facing after the training section solution maybe not satisfied if the given query is not matched with training section. Another major advantage of ANN is the memory distributed for large problem or component which is used within the network.

## **Application of ANN**

ANN is applied in various applications such as Social media Analytics, the medical field, image processing, pattern recognition, speech recognition, radar recognition, air traffic system, power supply system, cybersecurity, and many more. In the 1800s researcher builds an idea that is possible to make a technique or model that works like the human brain like thinking, self-learning, act according to behavior, and the ability of many other functions. A simple and basic ANN was developed between the 1960s and 1970s but fall due to lack of accuracy and efficiency, after that in 1986 improve the enhancement of ANN. The proposed system of this research which is used it consists of four layers: the first one is the sensor layer, the second layer is the object layer, the third phase of layer used the preprocessing layer, and the last layer is used in this proposed system is the application layer. The functionality of these mention layers is describing as sensor layers are also called the data acquisition layer and it consists of n-sensors for the purpose as input variables that are used to sense the environment and send the data to the object layer through a medium which is linked. After collecting data from sensors it will be saved in the form of raw material then data will be passed in the layer of preprocessing layer where analyze the data, mitigate the noise, normalize the data and make it in proper form for gaining the maximum accurate result because the collecting data may contain noise, irrelevant or missing information. After data arrange it will be exceeding into the next layer of the application layer. Where the application layer is divided into two main layers: the prediction layer and the performance evaluation layer.

The prediction layer is further divided into three layers of the input layer, hidden layer, and output layer and these three layers are perform on the base of Artificial Neural Network (ANN). ANN is a model whose function provides computational results with desire accuracy and efficiency. Three major qualities of ANN make it better than other computational methods and these qualities or advantages are ANN use parallel operation which faster than other computational methods those are used serial operation the reason behind this logic is that the parallelism methods has a high degree of sharing information or i/o signals because of distributed memory. Second is ANN learn from Data and store this information into the lookup table and third it uses nonlinear processing function to manage the complex nonlinear problems. ANN has very excellent properties such as self-learning, progressively in input to an output arranging, adaptively, and nonlinear problem that’s the reason ANN is mostly used for universal calculation in numerical standards.

Bitter et al. (2017) describe the importance of security not only in the pure IT department but also in those areas involved with the internet or computer. They defined that when we talk or concern with cyber security, we must accomplish three major qualities and these qualities are availability, confidentially and integrity and these qualities are satisfying with regulation, policies, and authorities. ANN successfully applied on a broad range of problem such as medical areas (heart, HIV aids, hepatitis, diabetes, etc.), finance (credit card, transactions, etc.), engineering (machine monitoring, robot, automobiles, etc.), Science (biology, chemistry, physics, nuclear physics, etc.).

Abiodun et al. (2018) describe the survey paper about the application of ANN which is used in real-world scenarios. These coming days ANN very popular and useful model for different disciplines like prediction, clustering, classification, pattern recognition, etc. This review study author furthermore presents the ANN application challenges, contribution, compare performance, evaluation methods.

Lek (2018) described that ANN is a powerful tool for prediction or solution of any problems and it has a very strong function and different types of algorithms or models also it shows universal and flexibility for any data. Prediction of modeling ANN is having a well-known approach use name ‘black box’ and this word black box means that all characters have an unknown situation, so these characters are identifying through training of ANN.

Potluri et al., (2017) described that Artificial Neural Network has a lot of advantages including requiring less formal statistical training to ANN for predict or finding the solution of a particular problem, ANN can point out all possible relation between predictor variables, there is a variety of algorithm and model are used in it for problem-solving and these algorithms and software/model and packages are easily available in the simple range, user-friendly, mostly graphical user interface (GUI), detect the non-linear relationship between dependent and independent variables, multiple algorithms are used to raining of an ANN, no need to know deep knowledge for developing/training of ANN just know about the basic structure of ANN model and data or parameters that can be adjusted with the model.

## **Data Mining in Intrusion Detection**

Dreossi et al. (2019) describe real-time data mining in intrusion detection systems (IDSs) and they have the most attention on this problem. Authors have mentioned that three types of issues that face in IDSs are accuracy, efficiency, and usability. The first problem is accuracy and it’s difficult to define how intrusion detects or performance or accuracy because naturally the data mining and IDSs have different definitions. Accuracy improves by analyzing the data and extracts data by using a data mining program and artificial anomalies make more improvement to detect the misuse or irregularity detection model by applying different algorithms and techniques. For improving efficiency in this paper use cost analysis and multiple models approach through this methodology proposed low cost and high efficiency they also present a distributed architecture with real-time for evaluating cost-sensitive. The third issue mentions in this paper was usability because IDSs which are based on data mining are more complex compare to a traditional system and this issue improving by an adaptive learning algorithm to develop the model construction and increasing the update for this purpose used unsupervised anomaly detection algorithm. In this research paper, they discuss and present an architecture which is consisting of detectors, model generation, sensors, data warehouse, and implementing this architecture improves the efficiency and scalability of the IDSs.

Tarik & Singh (2016) describe that how data mining is improving the intrusion detection in the field of cyber security because in present days the internet is more open and everyone is hanging on the internet for saving the information that’s why the security risks have been increased in the area of cybercrimes. In 2005 the pentagon publishes a paper which mentions that over 79000 effort to intrusion and 1300 are successful one so watching all these circumferences need to introduce new intrusion detection systems (IDSs) because older techniques are not enough for controlling the attack on cyber life. In this paper, the authors present the join of classifiers use with feature selection and multi-boosting techniques based on data mining for intrusion detection.

## **Go Deeper with Convolutions**

CNN's deep architecture has been effective in providing fine and excellent efficiency to trained models by learning patterns through raw images. A proposed deep neural network architecture is called Inception which attains the ILSVRC2014 (Image Net Large-Scale Visual Recognition Challenge 2014) assessment and the new architecture for identification. The basic goal of the design is to make better use of computing resources within the network. Through careful design, the depth and breadth of the network are increased when maintaining a computational and financial plan (Szegedy et al., 2015).

## **CNN Importance in Social Media Field**

Deep learning has supported the purpose of Computer Vision in recognizing and classifying images, and it is an important tool for automating tasks in daily lives. Object identification, classification, and segmentation have been developed using convolutional networks. Because of its ability to learn to represent data, the use of the convolutional neural network (CNN) on social media images has greatly helped social media analytics (Lecun et al., 2015).

## **Supervised Machine Learning Algorithm**

The (SML) Supervised Machine Learning is the exploration toward algorithms that cause after outwardly delivered cases to make overall assumptions, which then make calculations about upcoming cases. Most intelligent systems frequently use supervised classification. Algorithms like Linear Classifiers, (LR) Logistical Regression, Perceptron, NB Classifier, Support Vector Machine; Quadratic Classifiers, Boosting, Random Forest; and Neural networks algorithm are used where supervised machine learning deals with more classification (Osisanwo et al., 2017). Supervised Machine Learning methods need human beings to give essential inputs and outputs correspondingly, moreover, it gives estimation about the correctness of the calculation in the training procedure. Using the decision tree technique, the supervised ML algorithm is used to develop a prediction model using a labeled dataset to calculate social media analysis by the number of user responses (Muhammad et al., 2020).

## **Support Vector Machines**

The Support Vector Machines executes classification by creating an N-dimensional hyperplane that split up the data as classes. Concerning SVM, the analyst variable is called an element and the converted element is called a feature. Using the selection feature, the best and appropriate symbolic data is selected. Recounting a case in a set of features is known as a vector. Discovering the best hyperplane to split the clusters is one of the final objectives of this modeling. At the one side of the plane, there is a target variable whereas on the other side of the plane there is another group. Support vectors are the vectors that are near the hyperplane. The modules of support vector machines are nearly similar to classical multilayer perceptron neural networks (Oladipupo, 2016).

**CHAPTER 3**

1. **Project Design and Methodology**

Research Methodology is a method or computational techniques which are used for different purpose like process (observation, simulation, derived, experiment), data gathering, and analysis about a specific topic or problem. This research using three Binary Classification algorithms namely Multinomial Naïve bayes, Support Vector Machine Classifier and Multilayer Perceptron will attempt to identify and extract Fashion Related Data from Facebook. This will consist of experiments and data analysis.

Naïve Bayes is a famous probabilistic approach that is most popularly used for spam detection. In this research, the aim is to use the Naïve Bayes theorem for IE. The key to this is to make congruent the approach to IE with Spam detection. Spam detection works based on keywords. IE in this research will also work on the same principle. Based on specific keywords, Naïve Bayes can identify and thus help extract posts from Facebook that are Fashion related.

## **Information Extraction for social media**

In this exploration paper, the makers propose a design for Information Extraction (IE) from unstructured customer delivered substance-using online media (Badieh Habib Morgan and van Keulen, 2014). The design proposes answers for rout the IE challenges in this space like the short setting, the noisy meager substance, and the problematic substance. To overcome the challenges of standing up to IE from online media, State-Of-The-Art approaches ought to be changed by suit electronic media posts (Badieh Habib Morgan and van Keulen, 2014).

* + 1. **Named Entity Extraction**

The key sections and parts of the proposed framework are noisy substance filtering, named component extraction, named substance disambiguation, analysis circles, and weakness dealing with the generic methodology of information extraction from social media platform (Badieh Habib Morgan and van Keulen, 2014).

* + 1. **Named Entity Disambiguation**

Scientists regularly connect substances to Wikipedia articles or KB passages. For online media posts, here and there this is preposterous as a large number of the referenced substances can't be connected to Wikipedia articles or a KB passage (Badieh Habib Morgan and van Keulen, 2014). Notwithstanding, regularly clients have home pages or profiles on an online media network. Besides, celebrations and nearby occasions additionally generally have home pages addressing these occasions. This exploration proposes an open-world methodology for NED for tweets. Named elements are disambiguated by connecting them to a landing page or an informal community profile page if they don't have a Wikipedia article (Badieh Habib Morgan and van Keulen, 2014). Target (tweets rotating around the same occasion) are utilized to enhance the tweet setting and subsequently to improve the viability of tracking down the right substance page. Other metadata from client profiles could likewise be utilized to improve the disambiguation interaction (Badieh Habib Morgan and van Keulen, 2014).

* + 1. **Noisy Text Filtering**

It is needed to channel non-educational posts since there are a lot of tweets made each day and not all are valuable in this unique circumstance (Badieh Habib Morgan and van Keulen, 2014). Sifting should be possible dependent on area or language or different standards to try to keep just applicable posts that contain data about the space should be handled (Badieh Habib Morgan and van Keulen, 2014).

## **Lithium NLP: A System for Rich Information Extraction from Noisy User-Generated Text on social media**

Research done here displays the Lithium Natural Language Processing Framework – a tool obliged high throughput and language-agnostic framework for data extraction from loud client-produced text via online media (Bhargava, Spasojevic and Hu, 2017). Lithium NLP removes a rich arrangement of data including substances, subjects, hashtags, and notions from the text (Bhargava, Spasojevic and Hu, 2017).

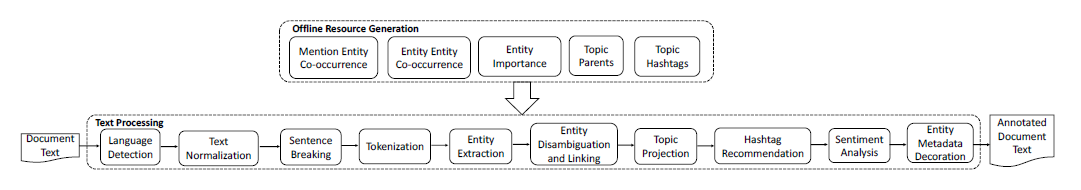


Figure 3.2: Lithium NLP Pipeline Overview (Bhargava, Spasojevic and Hu, 2017)

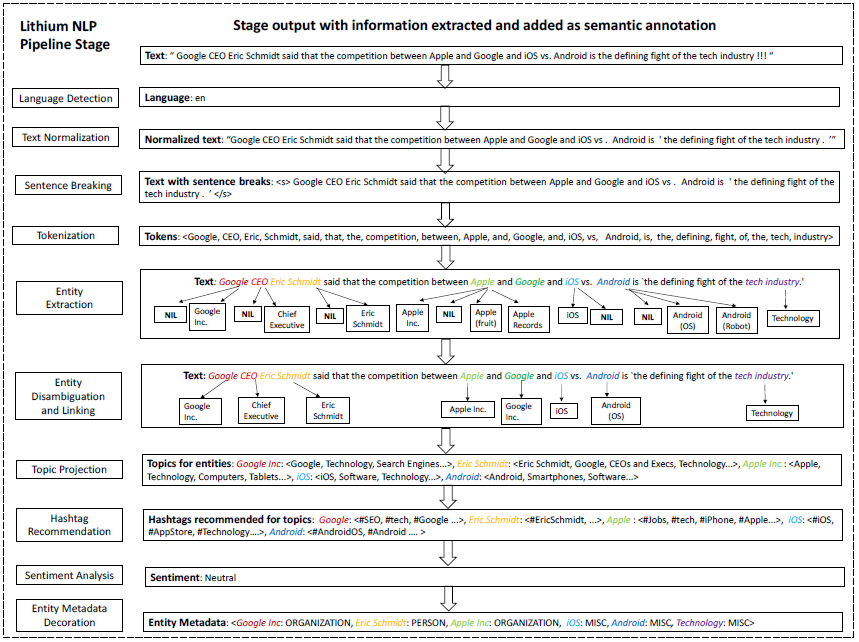


Figure 3.3: Example working of Lithium NLP (Bhargava, Spasojevic and Hu, 2017)

### 

## **Dataset**

The dataset for this research will be the Facebook Reactions dataset available at <https://github.com/minimaxir/interactive-facebook-reactions/tree/master/data>.

## **Mathematical Representation (Naïve Bayes):**

Naïve Bayes Theorem is represented by the following equation:

Where X and Y are two events, and D () represents the probability of the occurrence of the event. Naïve Bayes helps determine the occurrence of event X provided event Y has occurred.

Let,

* Event X be the occurrence of a specific keyword related to Fashion and,
* Event Y be the identification of a FB post as a Fashion related post.

The probability that a post would be Fashion related provided the keyword has already occurred is what Naïve Bayes will be providing. Since the Naïve Bayes theorem requires a set of keywords using which it would identify Fashion related posts from the chosen dataset, the methodology for extracting keywords is required.

For this research, the authors take data from four different domains for the domain analysis i.e., Social media, biomedical, chemical, and unstructured data. The reason for choosing these domains was their violent development rate helpful outcomes appearing daily from them.

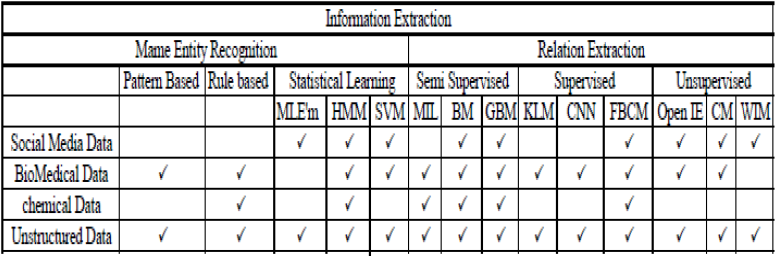


Figure 3.4: Popular Techniques for IE the Datasets used in this research (Mahboob Alam and Awan, 2018).

There are two major tasks in IE. Both have popular approaches that are widely used for research purposes.

## **Name Entity Recognition (NER)**

NER strives to locate and classify named entities from the content into predefined classes. The popular approaches used in this are as follows (Mahboob Alam and Awan, 2018):

## **Pattern-Based Method (PBM)**

This is the earliest of methods for IE (Mahboob Alam and Awan, 2018). Traditional IE methods use pattern recognition and word references to extract data from unlabeled data. PBM for IE is not easy to implement for social media. The reason behind this is that the data isn’t normalized. This is due to the risk of getting inaccurate results regarding patient diagnosis from the reports and patient data (Mahboob Alam and Awan, 2018). PBM isn’t used in the chemical domain also. This is due to the very large data dictionary of chemical compounds, atoms, and molecule names that isn’t possible for PBM to handle. Rule-based methods are used for data related to the chemical domain (Mahboob Alam and Awan, 2018).

## **Rule-based Method (RBM)**

This method has high congruency with Decision Trees (DT). The only difference is that in DTs overlapping is allowed which makes the training more robust (Mahboob Alam and Awan, 2018).

The RBM has been used for protein and gene entity recognition in the biomedical domain (Mahboob Alam and Awan, 2018). Using RBM makes it difficult to extract and label entities from unstructured data. But using word references i.e., seeds to kick-start that recognition process has worked in a few kinds of research (Mahboob Alam and Awan, 2018). RBM has also been used for question extraction and detection from Twitter microblogs. In the chemical domain, RBM has been used to identify entities as well as normalize the text while identifying the boundaries of the sentences in the text (Mahboob Alam and Awan, 2018).

## **Statistical Learning Method (SLM)**

SML refers to the representation of the empirical laws that govern Natural Language Training with a probability distribution. There are three popular approaches used in this method (Mahboob Alam and Awan, 2018):

## **Maximum Likelihood Estimation Method (MLEM)**

NER is easy for English because this language tends to be comprehendible, the rules are clear and the sentences are created by combining words (Mahboob Alam and Awan, 2018). But a language like Chinese is difficult since sentence making in Chinese is the total opposite of English (Mahboob Alam and Awan, 2018). MLEM has been used to successfully make sense of the unstructured Chinese data. MELM has been used for calculating probabilities in the biomedical domain. MLEM also helps increase the efficiency if MLEM is involved in the algorithm training process (Mahboob Alam and Awan, 2018).

MLEM cannot be used for any work in the chemical domain for the same reason PBM cannot be used (Mahboob Alam and Awan, 2018). The complexity of the data i.e., the equations, formulas is way too much. MLEM has been used for entity recognition from social media paired with the tag n-gram approach (Mahboob Alam and Awan, 2018).

## **Hidden Markov Model (HMM)**

HMM has been used successfully for NER in languages like Kannada that provide complex unstructured data. HMM was combined with the RBM approach for very impressive results (Mahboob Alam and Awan, 2018).

Social media-related datasets have a difficult time working with HMM due to irregular sentence structuring, message language, and irregular capitalization. But results have been achieved in this domain by combining HMM with n-gram for a hybrid approach N-gram Language Markov Model (NLMM) (Mahboob Alam and Awan, 2018). A much-needed success in the chemical domain regarding the extraction of information has been achieved by HMM. Linking chemical names of elements with their symbols i.e., the symbol for lead is PB (Mahboob Alam and Awan, 2018).

## **Support Vector Machines (SVM)**

SVMs use the power of linear algebra to create boundaries between different data classes. The intensity of separation is key in the results from SVM. This approach has been successfully utilized to extract individual information from the Swedish language like names, company location, and time. SVM has also been utilized for biomedical data NER (Mahboob Alam and Awan, 2018). The issue is dealing with social media data is that normal NER methods are created to deal with formal text properly written literature available in newspapers, etc. But social media doesn’t have this. Moreover, there is data in multiple languages on social media (Mahboob Alam and Awan, 2018). But SVM has outperformed HMM in this regard. The only disadvantage of SVM is that it is slow (Mahboob Alam and Awan, 2018).

## **Relation Extraction**

The second task in IE extracts the relationship between entities and the content of the data. The following popular approaches are used for this (Mahboob Alam and Awan, 2018):

## **Supervised Learning Models**

The goal here is to identify patterns and learn from a labeled dataset for prediction. The following approaches are popular in this section (Mahboob Alam and Awan, 2018):

## **Feature-Based Classification Method (FBCM)**

The difference between RBM and FBCM is the scope of data utilized. While RBM uses a focused, smaller scope of data for explaining the relations between entities and the context, FBCM uses a wide range of contextual data (Mahboob Alam and Awan, 2018). Traditional FBCM is unfit when it comes to social media IE. Therefore, relation extraction, a technique to extract the relations between linked entities in a text is used (Mahboob Alam and Awan, 2018).

The issue with FBCM is that occasionally information with unequivocal component vectors can't be addressed without any problem (Mahboob Alam and Awan, 2018). In these cases, the extraction of qualities is an exceptionally perplexing errand and prompts vectors of extremely high measurement, which thusly prompts issues of computation. Portion-based techniques endeavor to take care of this issue verifiably depend on scalar vector items are determined at extremely high Dimensional spaces with no sort of vector should be explicit (Mahboob Alam and Awan, 2018).

## **Semi-supervised Learning Methods**

Algorithms are trained on a dataset containing unlabeled and labeled data. This process is however expensive in terms of processing, time and data required. Using approaches that do not require much data is essential (Mahboob Alam and Awan, 2018).

## **Multi-Instance Learning (MIL)**

The preparation dataset for this methodology comes in areas, and each part is loaded up with numerous cases. MIL calculations discover events in the areas. At that point, those events are utilized for preparing. The events are described by the segments to which they have a place (Mahboob Alam and Awan, 2018).

## **Bootstrapping Method (BM)**

BM used to separate connections, has acquired impressive consideration lately (Mahboob Alam and Awan, 2018). These methodologies are built with a key suspicion that on the off chance that you a few words that relate with a particular goal in mind, phrases containing these words, these sorts of words probably make connections to communicate. Consequently, the sentences containing the word pair are utilized as preparing information for the connection extraction (Mahboob Alam and Awan, 2018).

## **Graph-Based Method (GBM)**

BM is notable for the extraction of connections, mostly since they require a tiny bit of measure of human appraisal. Charts can address complex connections among classes and occasions. An equivocal occurrence (Mahboob Alam and Awan, 2018); For instance, Usman Khawaja may be among the class of pilots and players. Today web-based media like Facebook beats the web. Scientists are chipping away at the investigation of online media for connection extraction from remarks, tweets just as posts (Mahboob Alam and Awan, 2018). In online media, the proposed execution of this strategy is a powerful spiral chart to adapt to the impediments of past representation procedures. The capacity to accurately get a handle on both semantic and syntactic designs in biomedical information ends up being dynamically essential and permits serious comprehension of logical papers and clinical information (Mahboob Alam and Awan, 2018).

## **Unsupervised Learning Methods**

In unaided realizing there is no such managed or organized information and we simply have input data. The goal is to find predictabilities at the event (Mahboob Alam and Awan, 2018).

## **Clustering Method (CM)**

Because of the huge number of connections between substances, it very well might be expensive to cover an adequately huge measure of preparing information to successfully stamp each sort of relationship in each new space of interest (Mahboob Alam and Awan, 2018). CM can be viewed as significant learning without managing the issue, like any such issue, attempt to discover a construction in an assortment of unlabeled information just as it is a vital unaided method. Substances are assembled by their absolute data. To utilize web-based media networks for the Semantic Web, a couple of surveys have dissected programmed connection extraction of online media (Mahboob Alam and Awan, 2018).

## **Open Information Extraction (Open IE)**

IE structures endeavor to remove the semantics of text in normal language associations (Mahboob Alam and Awan, 2018); in any case, most systems use oversaw express occasions of the relationship to learn and in this way most of the way by the ease of use of planning data (Mahboob Alam and Awan, 2018). Open IE framework depends absolutely upon the information substance and its etymological qualities. Much more particularly make advisers for getting these attributes of substance and from there on discrete relations (Mahboob Alam and Awan, 2018).

## **Wrapper Induction Method (WIM)**

This procedure is an extraction method that houses a course of action of extraction rules and program code needed to execute these principles (Mahboob Alam and Awan, 2018). WIM gets familiar with the covering naturally. Given a succession of preparing information, the acceptance calculation sorts out some way to remove a piece of covering objective information (Mahboob Alam and Awan, 2018).

**CHAPTER 4**

1. **Results and Discussion**

The project will be evaluated based on for this research are SVM, Multinomial Naïve Bayes and Multilayer perceptron. The coding section for this report is divided into two parts:

* **Data Pre-processing**
* **Machine Learning**

**Data Pre-processing**

In this section, the dataset is imported into the code using pandas and the dataset is then scanned for basic anomalies like duplicate instances and null values that need to be removed for better prediction results. Since there are multiple .csv files in the dataset of this research, all the data files are first to read and combined into one for easier management as shown below in 4.1 figure as a sample of the dataset dealt. This section is responsible for the data in the code. This section imports the dataset into the code, and analyzes it for any anomalies like duplicate datasets or null instances in the dataset and removes them. It also analyzes the dataset for any categorical features and encodes them using a Label Encoder which helps the dataset fit on the algorithm easily.

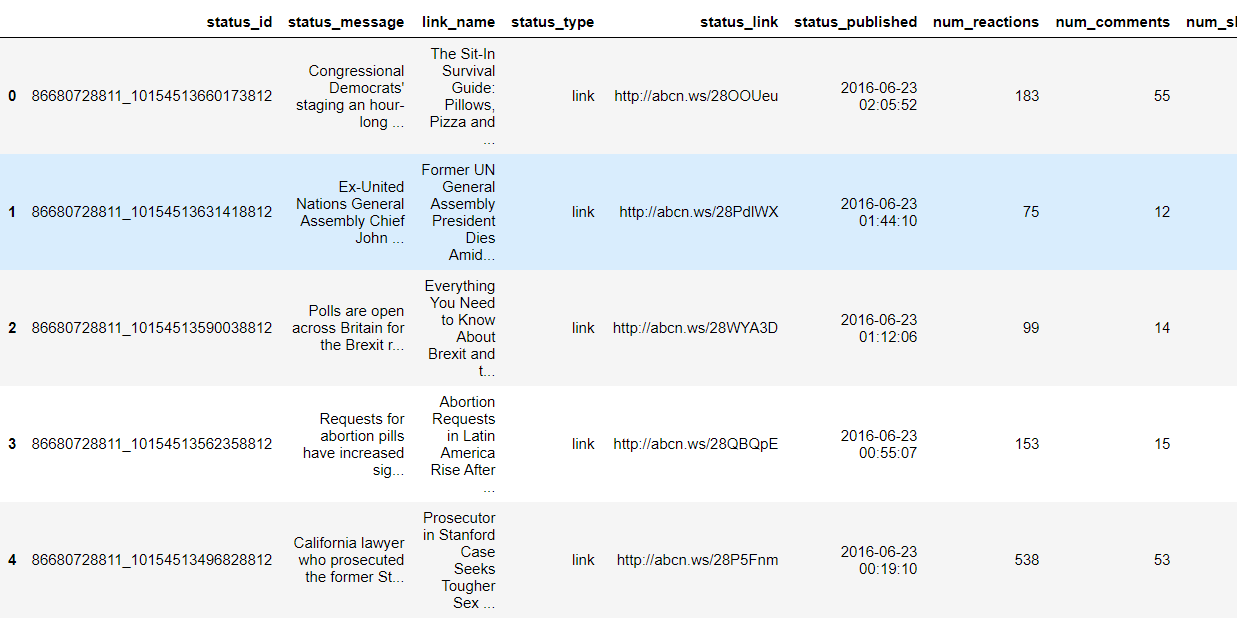


Figure 4.1: Dataset Sample Details

The first 5 instances of the dataset used in this research is shown in the figure below. The dataset has a total of 250 instances. Since the project requires an algorithm that can identify whether a facebook post is fashion related or not, this makes the problem a binary classification problem. The features in the dataset are as follows:

* link\_name: the name of the facebook post. This feature will be used to extracting the bag of words for Natural Language Processing.
* num\_reactions: total number of reactions on the post.
* num\_comments: total number of comments on the post.
* num\_shares: total number of shares for the post.
* num\_likes: total number of likes on the post.
* num\_loves: total number of love reacts on the post.
* num\_wows: total number of wow reacts on the post.
* num\_hahas: total number of haha reacts on the post.
* num\_sads: total number of sad reacts on the post.
* num\_angrys: total number of angry reacts on the post.

Label: denoting whether a post is fashion related or not using “n: for not fashion related” and “y: for fashion related”.

The dataset is divided using sklearn’s test-train-split function. The split ratio is 70:30 where 30% goes to test dataset and the rest 70 % goes to training set.

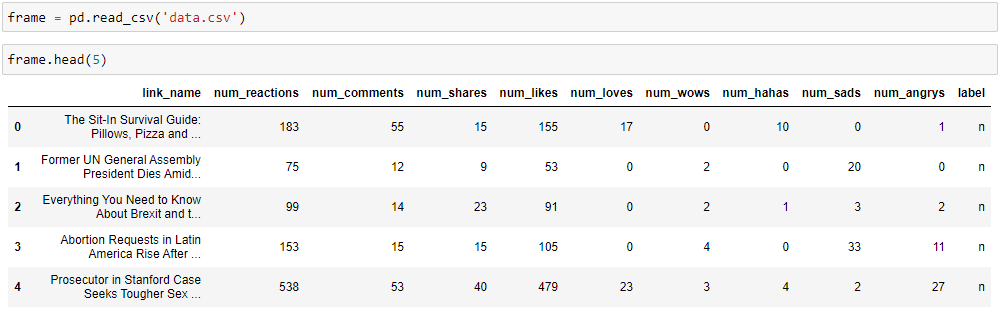


Figure 4.2: Importing the Dataset

The figure above shows the code responsible for importing the dataset into the code. This is verified by the first five instances of the dataset shown in the figure.

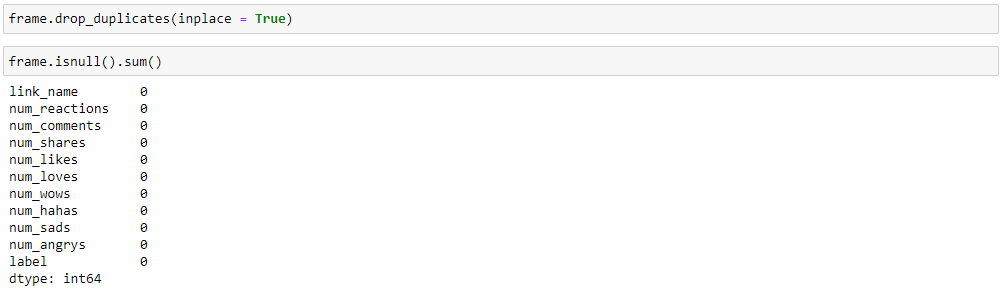


Figure 4.3: Remove the Null Instances from Dataset

The code above shows the duplicates and the null instances being removed from the dataset. The code below shows the statistical analysis on the dataset. Central tendency measures are used to understand the dataset. The metadata about the dataset is also displayed in the figure below that helps study about the features in the dataset.

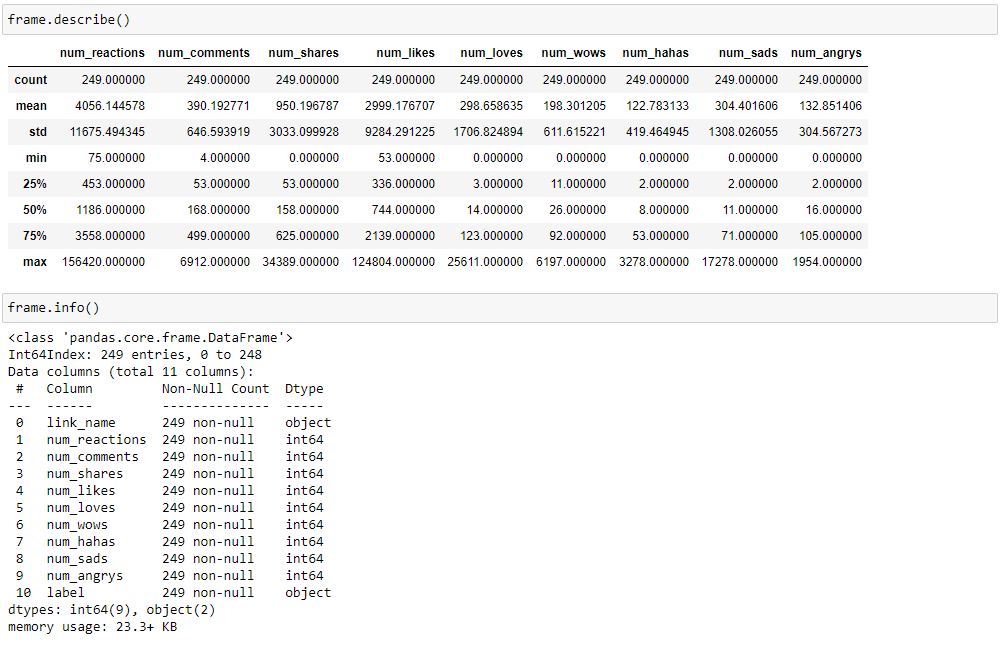


Figure 4.4: Metadata Details of Dataset

The code below is responsible for the Labe Encoding section of data pre-processing which is critical for training and testing the algorithm.

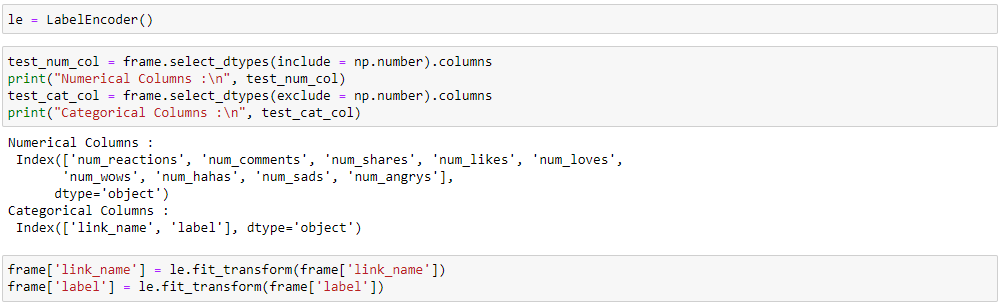


Figure 4.5: Labe Encoding

**Machine Learning**

The paradigm for this research in coding would be classification. The algorithms targeted for this research are SVM, Multinomial Naïve Bayes and Multilayer perceptron. The performance of the algorithms is measured using 2-fold cross validation and other classification performance metrics.

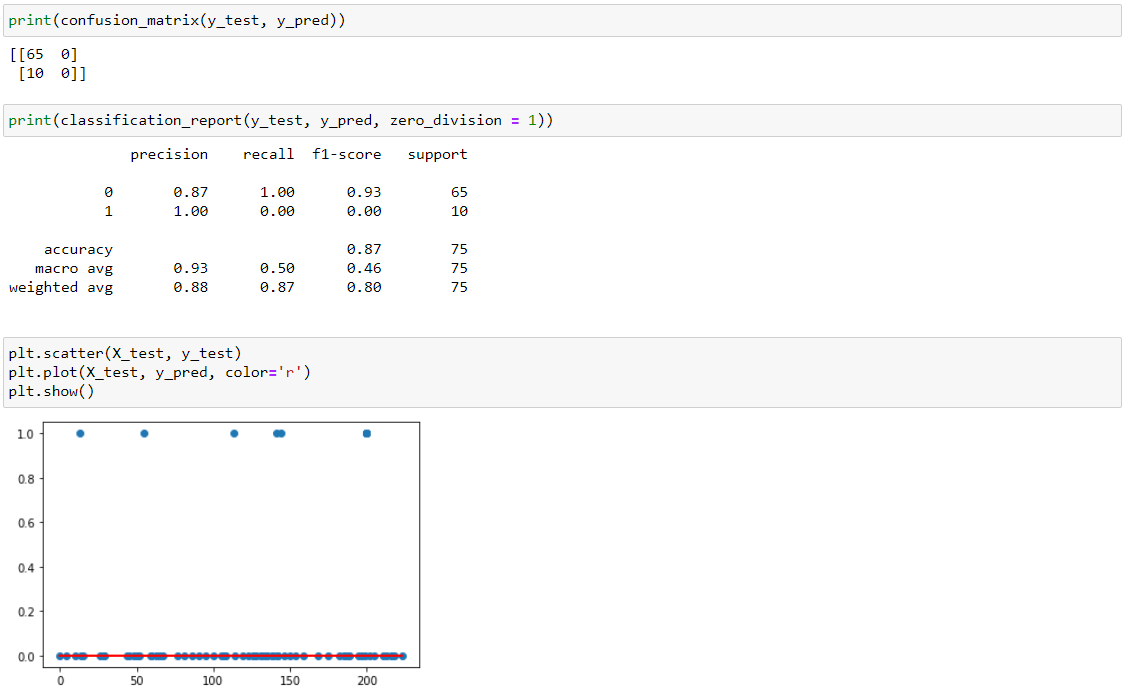


Figure 4.6: SVM Performa Evaluation

The above image shows the performance of the SVM algorithm. The blue scattered dots in the graph above is the dataset and the red line shows the fitting of the SVM algorithm on the dataset.

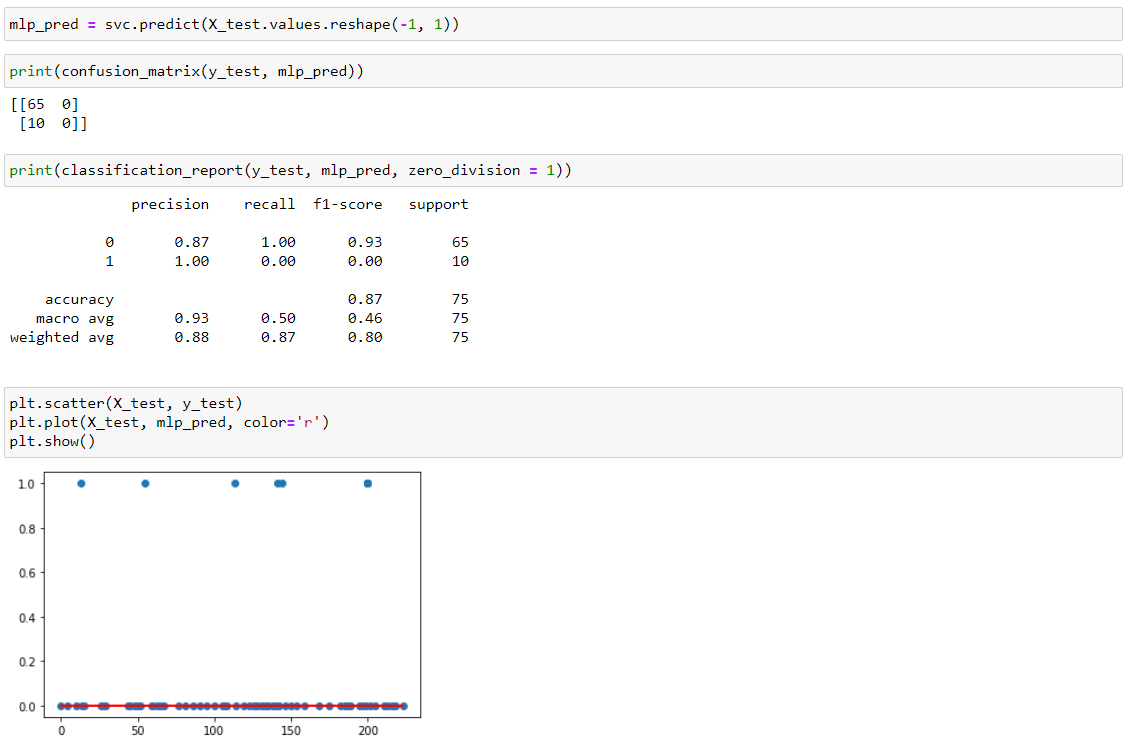


Figure 4.7: MLP Performa Evaluation

The above image shows the performance of the MLP algorithm. The blue scattered dots in the graph above is the dataset and the red line shows the fitting of the MLP algorithm on the dataset.

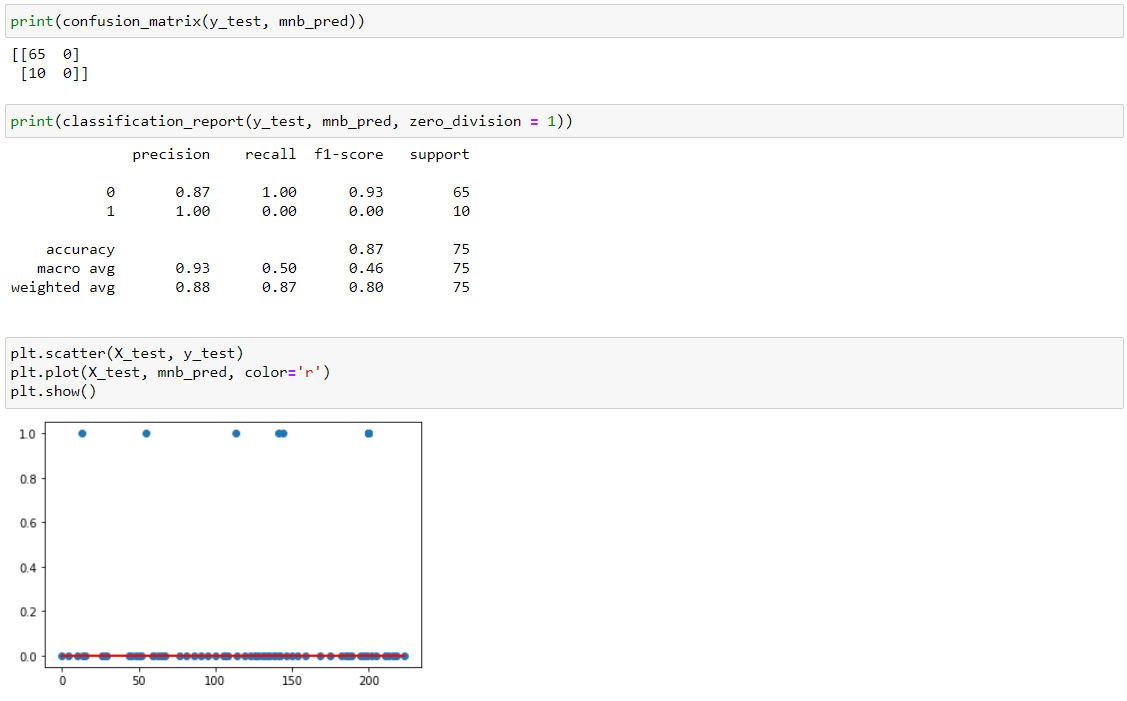


Figure 4.7: Naïve Bayes Performa Evaluation

The above image shows the performance of the Multinomial NB algorithm. The blue scattered dots in the graph above is the dataset and the red line shows the fitting of the Multinomial NB algorithm on the dataset.

**CHAPTER 5**

**Conclusion**

The purpose of this research extracting fashion-related news from Facebook enabled with machine learning to explore the ways social media influences fashion consumption of the male and female. The primary motivator for the usage of social media is to live in contact with friends and community. The experiment would be to provide the keywords related to Fashion news from Facebook for the Naïve Bayes using the TD-IDF approach. Then the Naïve Bayes will be trained using the selected dataset and the keywords that are collected from Facebook. The final performance will be measured using confusion matrices for evaluation.

**REFERENCES**

1. Michael, A., Armando, F., Rean, G., Anthony, D. J., & Katz, R. (2010). A view of cloud computing. Communications of the ACM.
2. Fan, W., & Gordon, M. D. (2014). The power of social media analytics. Communications of the ACM, 57(6), 74-81.
3. Gantz, J., & Reinsel, D. (2012). The digital universe in 2020: Big data, bigger digital shadows, and the biggest growth in the Far East. IDC iView: IDC Analyze the future, 2007(2012), 1-16.
4. Zeng, D., Chen, H., Lusch, R., & Li, S. H. (2010). Social media analytics and intelligence. IEEE Intelligent Systems, 25(6), 13- 16.
5. Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1), 1-167.
6. Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1), 1-167.
7. Karun, A. K., & Chitharanjan, K. (2013, April). A review on Hadoop—HDFS infrastructure extensions. In 2013 IEEE conference on information & communication technologies (pp. 132-137). IEEE.
8. Mangold, W. G., & Faulds, D. J. (2009). Social media: The new hybrid element of the promotion mix. Business Horizons, 52(4), 357-365.
9. Chen, C. P., & Zhang, C. Y. (2014). Dataintensive applications, challenges, techniques and technologies: A survey on Big Data. Information sciences, 275, 314-347.
10. Flood, M. D., Jagadish, H. V., & Raschid, L. (2016). Big data challenges and opportunities in financial stability monitoring. Banque de France, Financial Stability Review, 20, 129-42.
11. Kulkarni, A. P., & Khandewal, M. (2014). Survey on Hadoop and Introduction to YARN.
12. Coronel, C., & Morris, S. (2016). Database systems: design, implementation, & management. Cengage Learning.
13. Sowmya, R., & Suneetha, K. R. (2017, January). Data mining with big data. In 2017 11th International Conference on Intelligent Systems and Control (ISCO) (pp. 246-250). IEEE.
14. Basha, S. A. K., Basha, S. M., Vincent, D. R., & Rajput, D. S. (2019). Challenges in Storing and Processing Big Data Using Hadoop and Spark. In Deep Learning and Parallel Computing Environment for Bioengineering Systems (pp. 179-187). Academic Press.
15. Karpagam, N. S., Thrilokesh, G. K., Shanker, J., Harish, K., & Raja, M. (2017, February). Split reading of redundant datasets on data nodes using Hadoop framework. In 2017 International Conference on Algorithms, Methodology, Models and Applications in Emerging Technologies (ICAMMAET) (pp. 1-3). IEEE.
16. Jenhani, F., Gouider, M. S., & Said, L. B. (2019). Streaming Social Media Data Analysis for Events Extraction and Warehousing using Hadoop and Storm: Drug Abuse Case Study. Procedia Computer Science, 159, 1459-1467.
17. Amato, F., Moscato, V., Picariello, A., & Piccialli, F. (2019). SOS: A multimedia recommender System for Online Social networks. Future generation computer systems, 93, 914-923.
18. Chang, Y. C., Ku, C. H., & Chen, C. H. (2019). Social media analytics: Extracting and visualizing Hilton hotel ratings and reviews from TripAdvisor. International Journal of Information Management, 48, 263-279.
19. Suganya, S., & Selvamuthukumaran, S. (2018, March). Hadoop Distributed File System Security-A Review. In 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT) (pp. 1-5). IEEE.
20. Jimenez-Marquez, J. L., Gonzalez-Carrasco, I., Lopez-Cuadrado, J. L., & Ruiz-Mezcua, B. (2019). Towards a big data framework for analyzing social media content. International Journal of Information Management, 44, 1-12.
21. Berman, D. S., Buczak, A. L., Chavis, J. S., & Corbett, C. L. (2019). A survey of deep learning methods for cyber security. *Information*, *10*(4), 122.
22. Zhang, Y., Lee, W., & Huang, Y. A. (2003). Intrusion detection techniques for mobile wireless networks. *Wireless Networks*, *9*(5), 545-556.
23. Subroto, A., & Apriyana, A. (2019). Cyber risk prediction through social media big data analytics and statistical machine learning. *Journal of Big Data*, *6*(1),50.
24. Dlamini, M.T., Eloff, J.H., & Eloff, M.M. (2009). Information security: The moving target *computers & security*, *28*(3-4), 189-198
25. Meso, P., Ding, Y., & Xu, S. (2013). Applying protection motivation theory to information security training for college students. *Journal of Information Privacy and Security*, *9*(1),47-6
26. Craigen, D., Diakun-Thibault, N., & Purse, R. (2014). Defining cybersecurity. *Technology Innovation Management Review*, *4*(10).
27. Posey Garrison, C., & Ncube, M. (2011). A longitudinal analysis of data breaches. *Information Management & Computer Security*, *19*(4),216-230.
28. Gordon, L. A., & Loeb, M. P. (2006). Economic aspects of information security: An emerging field of research. *Information Systems Frontiers*, *8*(5),335-337.
29. Cazacu, M., & Cucu, C. (2019). Using the Activity Theory to Identify the Challenges of Designing E-learning Tools based on Machine Learning for Security Operations Centers. In *The International Scientific Conference eLearning and Software for Education* (Vol. 1, pp. 452-461). " Carol I" National Defense University.
30. Gordon, L. A., Loeb, M. P., Lucyshyn, W., & Zhou, L. (2014). Externalities and the magnitude of cyber security underinvestment by private sector firms: a modification of the Gordon-Loeb model. *Journal of Information Security*, *6*(01), 24.
31. Chan, P. K., & Lippmann, R. P. (2006). Machine learning for computer security. *Journal of Machine Learning Research*, *7*(Dec), 2669-2672.
32. Noor, U., Anwar, Z., Malik, A.W., Khan, S.,&Saleem, S.(2019).A machine learning framework for investigating data breaches based on semantic analysis of adversary’s attack patterns in threat intelligence repositories. *Future Generation Computer Systems*, *95*,467-487
33. Bhuyan, M. H., Bhattacharyya, D. K., & Kalita, J. K. (2013). Network anomaly detection: methods, systems and tools. *Ieee communications surveys & tutorials*, *16*(1), 303-336.
34. Garcia-Teodoro, P., Diaz-Verdejo, J., Maciá-Fernández, G., & Vázquez, E. (2009). Anomaly-based network intrusion detection: Techniques, systems and challenges. *computers & security*, *28*(1-2), 18-28.
35. O’connell, M. E., Arimatsu, L., & Wilmshurst, E. (2012, May). Cyber security and international law. In *International Law Meeting Summary, Chatham House* (pp. 5-7).
36. Jerome, O. U. (2012). *Cybersecurity Law & Regulation*. Wolf Legal.
37. Schjolberg, S. (2008). The history of global harmonization on cybercrime legislation–the road to geneva. *Journal of International Commercial Law and Technology*, *1*(12), 1-19.
38. Hemsley, K., & Fisher, R. (2018, March). A History of cyber incidents and threats involving industrial control systems. In *International Conference on Critical Infrastructure Protection* (pp. 215-242). Springer, Cham.
39. Syed, R., Khaver, A. A., & Yasin, M. (2019). Cyber Security: Where Does Pakistan Stand?.
40. Cherdantseva, Y., Burnap, P., Blyth, A., Eden, P., Jones, K., Soulsby, H., & Stoddart, K. (2016). A review of cyber security risk assessment methods for SCADA systems. *Computers & Security*, *56*, 1-27.
41. Peterson, C., & Rögnvaldsson, T. S. (1991). *An introduction to artificial neural networks* (No. LU-TP-91-23). CERN.
42. Bitter, C., Elizondo, D. A., & Watson, T. (2010, July). Application of artificial neural networks and related techniques to intrusion detection. In *The 2010 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE.
43. Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Mohamed, N. A., & Arshad, H. (2018). State-of-the-art in artificial neural network applications: A survey. *Heliyon*, *4*(11), e00938.
44. Lek, S., & Guégan, J. F. (1999). Artificial neural networks as a tool in ecological modeling, an introduction. *Ecological modeling*, *120*(2-3), 65-73.
45. Tu, J. V. (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of clinical epidemiology*, *49*(11), 1225-1231.
46. Potluri, S., Diedrich, C., & Sangala, G. K. R. (2017, September). Identifying false data injection attacks in industrial control systems using artificial neural networks. In *2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)* (pp. 1-8). IEEE.
47. Lee, W., Stolfo, S. J., Chan, P. K., Eskin, E., Fan, W., Miller, M., ... & Zhang, J. (2001, June). Real time data mining-based intrusion detection. In *Proceedings DARPA Information Survivability Conference and Exposition II. DISCEX'01* (Vol. 1, pp. 89-100). IEEE.
48. Dartigue, C., Jang, H. I., & Zeng, W. (2009, May). A new data-mining based approach for network intrusion detection. In *2009 Seventh Annual Communication Networks and Services Research Conference* (pp. 372-377). IEEE.
49. Dreossi, T., Donzé, A., & Seshia, S. A. (2019). Compositional falsification of cyber-physical systems with machine learning components. *Journal of Automated Reasoning*, *63*(4), 1031-1053.
50. Linda, O., Vollmer, T., & Manic, M. (2009, June). Neural network based intrusion detection system for critical infrastructures. In *2009 international joint conference on neural networks* (pp. 1827-1834). IEEE.
51. Tarik, M., & Singh, A. (2016)Intrusion Detection System using K2 Self Learning Algorithm and Open Attacking Platform. *International Journal of Scientific & Engineering Research, 6,* 543-548.
52. Liu, J., Xiao, Y., Li, S., Liang, W., & Chen, C. P. (2012). Cyber security and privacy issues in smart grids. *IEEE Communications Surveys & Tutorials*, *14*(4), 981-997.
53. Yan, Y., Qian, Y., Sharif, H., & Tipper, D. (2012). A survey on cyber security for smart grid communications. *IEEE Communications Surveys & Tutorials*, *14*(4), 998-1010.
54. Mathur, M., 2018. Where is the Security Blanket? Developing Social Media Marketing Capability as a Shield from Perceived Cybersecurity Risk. Journal of Promotion Management, 25(2), pp.200-224.
55. Konyeha, S., 2020. Exploring Cybersecurity Threats in Digital Marketing. NIPES Journal of Science and Technology Research, 2(3), p.12.
56. Benz, M. and Chatterjee, D., 2020. Calculated risk? A cybersecurity evaluation tool for SMEs. Business Horizons, 63(4), pp.531-540.
57. Suh, T. and Chow, T., 2021. Developing a digital marketing tool for ethnic ventures' mixed business model and market-shaping: A design scientific approach of web demographics. Industrial Marketing Management, 93, pp.10-21.
58. Baird, C. and Parasnis, G. (2011). From social media to social customer relationship management.Strategy \& Leadership, 39(5), pp.30--37.
59. Berthon, P., Pitt, L., Plangger, K. and Shapiro, D. (2012). Marketing meets Web 2.0, social media, and creative consumers: Implications for international marketing strategy. Business Horizons, 55(3), pp.261--271.
60. Beuker, R. and Abbing, E. (2010). Two Faces of Social Media: Brand Communication and Brand Research. Design Management Review, 21(1), pp.54--60.
61. Blaxter, L., Hughes, C. and Tight, M. (2010). How to research. 1st ed. Maidenhead: Open University Press/McGraw-Hill Education.
62. Brodie, R., Ilic, A., Juric, B. and Hollebeek, L. (2013). Consumer engagement in a virtual brand community: An exploratory analysis. Journal of Business Research, 66(1), pp.105--114.
63. Bryman, A. (2012). Social research methods. 4th ed. Oxford: Oxford University Press.
64. Chen, Y., Fay, S. and Wang, Q. (2011). The role of marketing in social media: How online consumer reviews evolve. Journal of Interactive Marketing, 25(2), pp.85--94.
65. Costa, R. (2013). The Effects of Online Brand Communities on Brand Equity in the Luxury Fashion Industry. International Journal of Engineering Business Management Special Issue on Innovations in Fashion Industry. [online] Available at: http://cdn.intechopen.com/pdfs-wm/45560.pdf [Accessed 29 May. 2014].
66. Dellarocas, C. (2003). The digitization of word of mouth: Promise and challenges of online feedback mechanisms. Management science, 49(10), pp.1407--1424.
67. De Vries, L., Gensler, S. and Leeflang, P. (2012). Popularity of brand posts on brand fan pages: an investigation of the effects of social media marketing. Journal of Interactive Marketing, 26(2), pp.83--91.
68. Doç, Y. and Ulusu, Y. (2010) ‘Determinant factors of time spent on Facebook: brand community engagement and usage types’ Journal of Yaşar University, 18(5) 2949-2957.
69. Socialbakers.com. Inspiration from the Fashion Industry On Social Media. [on line]. socialbakers.com, [Accessed 30 July. 2014].
70. "Finding Success with Social Media", 2010, Marketing Weekly News, , pp. 80.
71. Friege, C., Gensler, S., Rangaswamy, A., Skiera, B., Malthouse, E.C., Lobschat, L. & Hennig-Thurau, T.( 2010). "The impact of new media on customer relationships", Journal of service research, vol. 13, no. 3, pp. 311-330.
72. Gummerus, J., Liljander, V., Weman, E. and Pihlstr\"om, M. (2012). Customer engagement in a Facebook brand community. Management Research Review, 35(9), pp.857--877.
73. Habibi, M., Laroche, M. and Richard, M. (2014). The roles of brand community and community engagement in building brand trust on social media. Computers in Human Behavior, 37, pp.152--161.
74. Heinonen, K. (2011). Consumer activity in social media: Managerial approaches to consumers' social media behavior. Journal of Consumer Behaviour, 10(6), pp.356--364.
75. Heinonen, K. (n.d.). Social Media Activities: Understanding What Consumers Do in Social Media.
76. Hutton, G. and Fosdick, M. (2011). The globalization of social media consumer relationships with brands evolve in the digital space. Journal of advertising research, 51(4), pp.566 - 570.
77. Kim, A. and Ko, E. (2012). Do social media marketing activities enhance customer equity? An empirical study of luxury fashion brand. Journal of Business Research, 65(10), pp.1480--1486.
78. Kozinets, R.V. (1999). "E-tribalized marketing?: the strategic implications of virtual communities of consumption", European Management Journal, vol. 17, no. 3, pp. 252-264.
79. Krishnamurthy, S. and Dou, W. (2008). Note from special issue editors: advertising with user-generated content: a framework and research agenda. Journal of Interactive Advertising, 8(2), pp.1--4.
80. Mohr, I. (2013). The Impact of Social Media on the Fashion Industry. Journal of Applied Business and Economics, 15(2), pp.17--22.
81. Moran, E. 2010. Marketing in a hyper-social world – The tribalization of business study and characteristics of successful online communities. Journal of Advertising Research,: 232–239. doi:10.2501/S0021849910091397
82. Muk, A., Chung, C. and Kim, J. (2014). A Cross-National Study of the Influence of Individualism and Collectivism on Liking Brand Pages. Journal of International Consumer Marketing, 26(2), pp.122--137.
83. Digital Firefly Marketing, (2014). Social Media Trends: Why do people follow a brand?. [online] Available at: https://digitalfireflymarketing.com/social-mediatrends-why-do-people-follow-brand [Accessed 27 Jul. 2014].
84. Badieh Habib Morgan, M. and van Keulen, M., 2014. Information Extraction for Social Media. Proceedings of the Third Workshop on Semantic Web and Information Extraction,.
85. Bhargava, P., Spasojevic, N. and Hu, G., 2017. Lithium NLP: A System for Rich Information Extraction from Noisy User Generated Text on Social Media. Proceedings of the 3rd Workshop on Noisy User-generated Text,.
86. Mahboob Alam, T. and Awan, M., 2018. Domain Analysis of Information Extraction Techniques. INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY SCIENCES AND ENGINEERING, 9(6).