

Geospatial Data Analysis for Mapping Customers Complaints (Case Study: First Media Users Using Social Media of Twitter)

Ulfah Hayati Baros
{upebaros8@gmail.com}

Universitas Gunadarma, Indonesia

Abstract. The role of information technology in people's lives has become a necessity that cannot be separated from various aspects of life such as in the fields of communication, health, education, security, business activities, and others. The growing need for cellular data connections has triggered the growth of the internet service provider business market. PT First Media Tbk is one of the largest internet service providers in Indonesia. The growing growth of users, of course, cannot be separated from the disruption in internet use using First Media. First Media also experienced customer complaints from various customer services. In the use of social media, Twitter is one of the social media used by First Media in delivering and updating internet service information. Users can submit complaints on First Media's Twitter account, @FirstMediaCaresCares. Services on Twitter social media can be used to produce more valuable information through the Twitter Data Analytics process and utilize geospatial data on Twitter social media. The data obtained in the form of words that come from customer posts on Twitter social media are the basis for collecting customer data on complaints from the use of First Media which is then mapped to form the coordinates of Twitter users. The coordinate point is obtained from sharing location by Twitter users. The extraction and analysis process is carried out in four stages, namely: the process of data retrieval (crawling), storage (storing), analysis (analyzing), and visualization (visualizing). Extraction and analysis of Twitter's geospatial data on First Media's customers who report internet disruption complaints are carried out in order to find out which customers in which areas have frequent internet disruptions.

Keywords: data analysis; social media analytics; twitter data analytic; geospatial

1 Introduction

The role of information technology in people's lives has become a necessity that cannot be separated from various aspects of life such as in the fields of communication, health, education, security, business activities, and others. The existence of information technology can support a company in conducting business processes effectively and efficiently. One of the technologies that is growing in today's era is the internet. The need for an increasingly large cellular data connection has triggered the growth of the internet service provider business market. Provider companies in Indonesia are competing to attract customers, such as PT First Media Tbk which is one of the leading provider companies in Indonesia [1].

The growing user growth certainly can't be separated from the disruption in internet usage using First Media. First Media also have experienced customer complaints from various customer services. Customers can report complaints through various First Media services such as call centers, email and social media. Services such as call centers and emails can only be accessed internally by the company to find out customer reports, while social media services can be accessed and known by the public, so that social media can be used as a forum to obtain information on customer complaints openly. Twitter is one of the social media used by First Media in conveying and updating internet service information. Users can submit complaints on the First Media Twitter account, namely @FirstMediaCaresCares. This service on Twitter can be used to generate more valuable information through the Twitter Data Analytics process and utilize geospatial data on Twitter.

This Twitter data analysis applies model building for prediction and processing data through words that appear in the microblog data set (Twitter). When extracting and analyzing geospatial data on Twitter, knowledge and frameworks of social media analytics (SMA) are required. The extraction and analysis process is carried out through four stages, namely: crawling, storing, analyzing, and visualizing. In this research, the extraction and analysis of Twitter geospatial data was carried out on First Media customers who reported complaints of internet disturbances so that they could find out which customers in the area often experienced internet disturbances.

2 Methodology

This research focuses on extracting geospatial data through Twitter social media from every tweet sent based on the user's location. Each region is identified by latitude and longitude coordinates obtained from google maps. The research method is descriptive to describe the location of each user's tweet that has been sent to the @FirstMediaCares account. The data used in this study is data on the Twitter account @FirstMediaCares. The data that is withdrawn from the @FirstMediaCares account is carried out for a duration of 7 days from 28 March to 03 April 2021. The data withdrawal is related to user complaints against the use of First Media which will be mapped based on the location of the user. The research stages are shown in Figure 1.

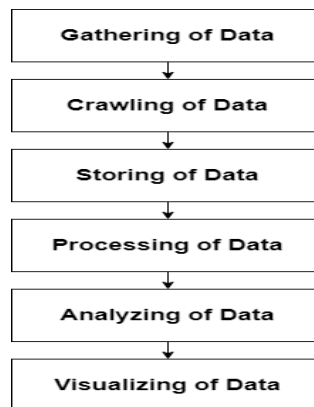


Fig. 1. Stage of Research.

a. Gathering of Data

The data collected is data from Twitter social media obtained through user tweets sent to the @FirstMediaCares account and posts from the @FirstMediaCares account. The data is taken based on the availability of data contained in the Twitter database on the @FirstMediaCares account for a period of 7 days from March 28 to April 03, 2021. The data obtained are real.

b. Crawling of Data

The data retrieval process uses the JavaScript programming language. Retrieval of data using twitter streaming API. The data drawn is public data where the data drawn is divided into 2 categories, that are:

- Crawling of data on user tweets where the user replies, retweets, and mentions the @FirstMediaCares account.
 - Crawling of data on tweets made by the @FirstMediaCares account.
- c. Storing of Data

The data that has been found through the data crawling process is stored in a file in the form of JSON, XML, CSV, XLS or into a database. The data that has been successfully stored is then analyzed to get clean data so that the data can be processed properly.

d. Processing of Data

Crawling data that has been saved is raw data and can't be used for analysis purposes. The raw data is processed by cleaning, filtering and adding the required fields

e. Analyzing of Data

Data analysis is an important part, especially in data processing. This data processing is carried out so that the data obtained does not contain empty or null files, if there is such a file then the data analysis does not get results. After the data has been analyzed, then from the results of the analysis, the data visualization stage can be carried out

f. Visualizing of Data

Visualization is a way to convert data into a visual or tabular format so that the characteristics of the data and the relationships among data items or attributes can be analyzed or reported. Data visualization is one of the best and most interesting techniques in terms of exploring data. Data visualization is intended so that the data generated from the analysis process looks more attractive and can be more easily understood as information.

3 Result and Discussion

This chapter will discuss the results of implementation and trials carried out after going through the stages of the previous research method. At this stage, a trial of retrieval of tweet data is carried out through the Twitter streaming API. Tweet data taken is data that is Public Streams. In collecting data via Twitter, it is necessary to create a Twitter Developer account first to be able to legally withdraw data with permission from Twitter itself. The results of each data processing are processed using Visual Studio Code and then visualized in the form of tables and graphs using the Tableau platform.

a. Creation of Developer Twitter Account

Before collecting data from twitter we need to create a developer account first. The function of creating a twitter account for this developer is to collect tweet data from firstmedia consumers through the Twitter API. Twitter API has a security mechanism by using identities called Consumer Keys and Access Tokens. This identity is needed because all things related to the Twitter API in any programming must use the key and token to be processed using JavaScript. The consumer key are shown in Figure 2.

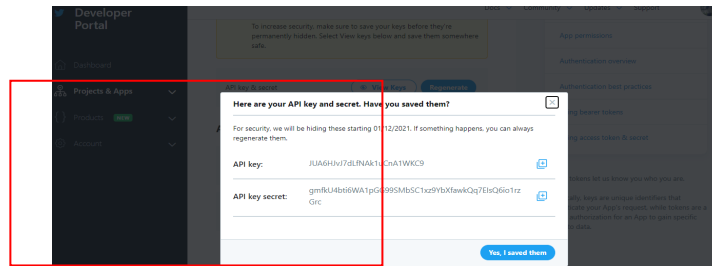


Fig. 2. Consumer key on Developer Twitter Account.

b. Backend Project Creation

The creation of a backend project is carried out for crawling the existing data on Twitter for every tweet of customer complaints sent to the @FirstMediaCares account and configuring the database which will be implemented in an interface so that it will be known whether the system that has been created is appropriate. Making a backend project consists of installing libraries, frameworks, package names, environments, servers, and database configurations.

c. Twitter API Call

In the data crawling process to display information about a tweet determined by the consumer key and access token that has been obtained previously, a variable that calls the Twitter API is created, if successful, it will get a response in the form of a tweet, if it fails, it will get an error response.

```

1 // Deklarasi memanggil library, fungsi & variable
2 const axios = require("axios");
3 const util = require("../util/util");
4
5 // Variable base url API twitter
6 const BaseURL = "https://api.twitter.com/2/tweets";
7 // Konfigurasi header yang dikirim ke twitter
8 const config = {
9   headers: {
10    Authorization:
11    "Bearer AAAAAAAAAAAAAAAAAAAAF3NjwEAAAAAQ1n3vpP51M8YR2F11o71tyfmC6SvoX3D0FHicHCcAe1dpUESAgTjtpcuq49R4P1s
12   },
13 };
14 // Variable untuk response
15 let response = {
16   success: false,
17   result: null,
18   error: null,
19 };

```

Fig. 3. API calls on Twitter.

d. Cleansing and Filtering Data

Furthermore, after receiving a response in the form of a tweet, filtering and cleansing of the tweet was carried out according to the complaint category, namely network, billing and cable tv.

```

// Deklarasi Variable untuk tanggal start time crawling data
let startDate = new Date();
startDate = new Date(startDate.setDate(startDate.getDate() - 7));
startDate = new Date(startDate.setMinutes(startDate.getMinutes() + 5));
startDate = startDate.toISOString();

// Looping untuk mendapat data dari API Twitter
do {
// Hit Service API untuk get Data Twitter, mengirinkan tanggal yang sudah di deklarasikan & token
const csvData = await service.get(startDate, nextToken);

// Kondisi apabila terdapat error pada service API
if (csvData.success === false) {
// Apabila terjadi error membuat object & flag error dan keluar dari looping
flag = false;
flagError = true;
response.success = false;
response.error = {
status: csvData.error.status,
error: csvData.error.error,
message: csvData.error.message,
};
} else {
// Pengecekan token dari hasil service API
if (csvData.result.meta.next_token) {
nextToken = csvData.result.meta.next_token;
} else {
nextToken = null;
flag = false; // Apabila false, perulangan berakhir setelah proses selesai.
}
}

// Seleksi kriteria/kategori tweet
if (
text.includes("lemot") ||
text.includes("lambat") ||
text.includes("stabil") ||
text.includes("modem") ||
text.includes("gangguan") ||
text.includes("mati") ||
text.includes("sinyal") ||
text.includes("lama") ||
text.includes("offline") ||
text.includes("trouble") ||
text.includes("trouble") ||
text.includes("loss") ||
text.includes("packet") ||
text.includes("jaringan")
) {
dataElement.kategori = "Jaringan";
} else if (text.includes("tagihan") || text.includes("bayan")) {
dataElement.kategori = "Tagihan";
} else if (
text.includes("tv") ||
text.includes("homecable") ||
text.includes("channel")
) {
dataElement.kategori = "TV Kabel";
} else {
skip = true;
}
}
}

```

Fig. 4. Cleansing and Filtering Data.

Variable declaration for crawling data with a period of 7 days starting from March 28 to 03 April 2021 with a total of 594 data as shown in Figure 4 above. If the content of the tweet matches the category that has been created, it will be executed directly into the database, if it does not match the category, such as the absence of information about the location of the tweet and tweet data that is not related to the complaint, the tweet is not used.

```

// Pengecekan data
if (data.length > 0) {
// Response data apabila terdapat data dari database
return res.status(200).json({ status: 200, message: "Success", data });
} else {
// Response data apabila tidak terdapat data dari database
return res.status(404).json({ status: 404, message: "Data Not Found" });
}
} catch (error) {
// Response Error
console.error(error);
return res.status(500).json(error);
}
};

```

Fig. 5. Data Checking.

When checking the data as shown in Figure 5, if the data is still in the main Twitter database with a range according to the token that has been obtained, it will display all tweet data according to the category with a period of 7 days, if it is not in the Twitter database, it will displays the error message "Data Not Found".

e. Convert Twitter Data to Excel

When the data crawling has been successfully carried out, then the next step is to convert the data into an excel table so that it is easy to visualize the data. Creating a table in .xls format containing the date, user ID, tweet, latitude, longitude, place and category as shown in Figure 6.

```
// Pembuatan Excel dengan Excel JS
const workbook = new ExcelJS.Workbook();
workbook.creator = "Ulfah Hayati Baros";
workbook.created = new Date();
const worksheet = workbook.addWorksheet("Data Twitter", {
  pageSetup: { paperSize: 9, orientation: "potrait" },
});
worksheet.columns = [
  { header: "Tanggal", key: "tanggal" },
  { header: "ID", key: "id" },
  { header: "Tweet", key: "tweet" },
  { header: "Latitude", key: "latitude" },
  { header: "Longitude", key: "longitude" },
  { header: "Place", key: "place" },
  { header: "Kategori", key: "kategori" },
];

// Menambahkan data ke dalam Excel
worksheet.addRow(data);

worksheet.addRow([
  {
    tanggal: "Kategori:",
    id: `Jaringan = ${countJaringan}`,
  },
  {
    id: `TV Kabel = ${countTv}`,
  },
  {
    id: `Tagihan = ${countTagihan}`,
  },
]);
```

Fig. 6. Convert Twitter Data to Excel.

f. Data Interface Implementation

In the implementation of the data in the form of a website page interface using localhost, it displays data that has been obtained previously in the form of a table containing the date, ID, Tweet, Latitude, Longitude, Place and Category. On the website display, functions such as create data are also provided to retrieve data automatically based on 7 days ago which was calculated when data crawling was carried out. Then there is a data filter function that can enter the date of the data you want to display with a maximum period of 7 days, then there is a function to download the data in .csv or .xls format which makes data visualization easier.

Tanggal	ID	Tweet	Latitude	Longitude	Place	Kategori
2021-10-20T14:27:44.000Z	1370107192436200	@HafidKurniaku Haha banget program jaringannya udah selesai di 40 Puncak aja udah di internetkan wad	-1.57768	107.10342	Bogor, Indonesia	Jaringan
2021-10-20T14:26:53.000Z	1370106447750000	@HafidKurniaku dan 195633327 mau pin 20 sampai 28 mau pin 20 itu wad internet dan udah 7x. bisa on the range 28' need dan mau lg sampai. Boleh pgg? hepi hepi!	1.64558	107.42553	Diponegoro, Indonesia	Jaringan
2021-10-20T14:26:00.000Z	1370106291915700	@HafidKurniaku ini internet not true, karena ckt dari daku. DM app ya bisa digunakan di sublaya	-1.55248	107.14390	Gede, Indonesia	Jaringan
2021-10-20T14:25:10.000Z	1370106145880000	@HafidKurniaku mules, bingung apa program di internet apa mules ni	-1.67762	107.28589	Cirebon, Indonesia	Jaringan
2021-10-20T14:24:30.000Z	1370106103200000	@HafidKurniaku mules, bingung apa program di internet apa mules ni	-1.58196	107.22983	Ikarya, Indonesia	Jaringan
2021-10-20T14:23:50.000Z	1370106079840000	@HafidKurniaku @pryptoblog ya all hepi hepi bisa internetkan, udah koneksi di ping	-1.67703	107.45557	Bogor, Indonesia	Jaringan
2021-10-20T14:23:10.000Z	1370106052000000	@HafidKurniaku bingung apa mau main di jaringan tempo dulu? / noten hopidolnya	1.56020	107.24575	Semarang, Indonesia	Jaringan
2021-10-20T14:22:30.000Z	1370106046700000	W @HafidKurniaku tempo internet mail? / noten hopidolnya	-1.52029	106.89536	Cirebon, Indonesia	Jaringan
2021-10-20T14:21:50.000Z	1370106033000000	@HafidKurniaku dan kemana pgn demo off? / noten hopidolnya	1.42937	107.24568	Semarang, Indonesia	Jaringan
2021-10-20T14:21:10.000Z	1370106024800000	@HafidKurniaku sama demo error no. tidak ada signal terdeteksi. internetnya juga error	1.57610	107.38729	Cirebon, Indonesia	Tv Kabel

Fig. 7. Data Display on Interface

g. Data Visualization in Network Category

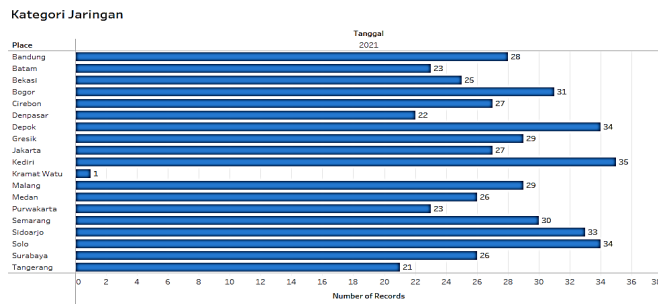


Fig. 7. Graph of Total Customer Complaints for Network Category March 28 to April 03, 2021.

Based on the graph in Figure 7, in the network category from March 28 to April 03, 2021 for 7 days with a total of 510 data, the most customer complaint data obtained is in the City of Kediri with 35 data.

h. Data Visualization in Billing Category

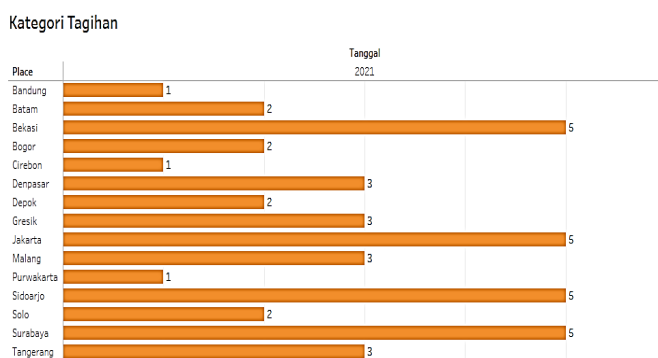


Fig. 8. Graph of Total Customer Complaints for Billing Category March 28 to April 03, 2021.

Based on the graph in Figure 8, in the billing category from March 28 to April 03, 2021 for 7 days with a total of 42 data, the highest customer complaint data was obtained in the cities of Bekasi, Jakarta, Sidoarjo, and Surabaya, each with 5 data.

i. Data Visualization in Cable TV Category

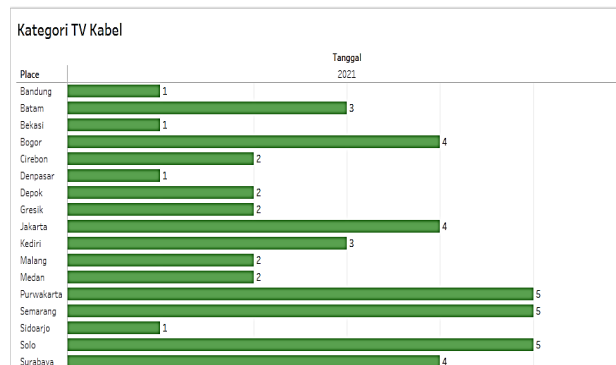


Fig. 9. Graph of Total Customer Complaints for Cable TV Category March 28 to April 03, 2021.

Based on the graph in Figure 9, in the cable tv category from March 28 to April 03, 2021 for 7 days with a total of 42 data, the most data obtained from customer complaints are in the cities of Purwakarta, Semarang and Solo, each with 5 data.

j. Overall Total Data Visualization

Visualization of FirstMedia's customer complaint data at this stage is carried out in its entirety where every tweet from all categories is in which location has the most complaint data for 7 days starting from March 28 to April 03, 2021 as shown in Figure 10.

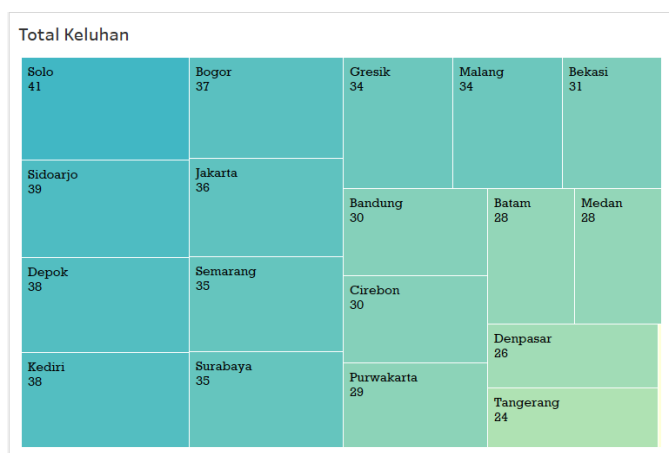


Fig. 10. Total Customer Complaints 28 March to 03 April 2021.

Based on the treemaps in Figure 10, overall from March 28 to April 03, 2021 for 7 days with a total of 594 data, the highest customer complaint data was obtained in the City of Solo with 41 data. Complaints related to the network as many as 34 data, 2 data bills and 5 data cable tv.

4 Conclusion

Based on the results of the analysis that has been carried out using descriptive analysis methods on the tweet data of complaints against FirstMedia users obtained through mentions on the @FirstMediaCares twitter account, the following conclusions are obtained:

- Geospatial data analysis for mapping First Media customer complaints which is processed based on data from social media twitter provides information about the location of customers with any type of complaint being reported on the @ FirstMediaCares twitter account which is a social media-based service for First Media customers where First Media can evaluate and respond to disturbances experienced by customers based on the location and types of disturbances that customers often face as seen from customer complaint tweets.
- Tweet data related to customer complaints regarding the use of First Media services for 7 days from 28 March to 03 April 2021 was successfully carried out with a crawling cleansing and filtering process using the Twitter Streaming API with a total of 594 data.
- The results of geospatial data visualization, obtained information about customer complaints with the highest number based on network categories, there are Kediri City as many as 35 complaints data, the Billing category there are Bekasi, Jakarta and Sidoarjo as many as 5 complaints data, cable tv categories there are Purwakarta, Semarang and Solo as many as 5 complaints data, and for the whole of the most total complaint data, there are 41 complaints in Solo City.
- Location data from each customer can only be obtained if the customer activates location from the customer's device.

Geospatial data analysis for mapping First Media customer complaints through Twitter social media can be further developed with other social media facilities that can be compared in terms of the number of customer accounts on each social media and also how many customer complaints reports. There are limitations that occur because the privacy of customers who want or not to activate

the location makes geospatial-based research less than optimal, it is better for further research to use other methods so that it can support data accuracy with more perfect geospatial.

References

- [1] Abbasi, A, Hassan, A, & Dhar, M 2014, 'Benchmarking twitter sentiment analysis tools'. Journal of In LREC, vol. 2, no. 12, hh. 823-829.
- [2] B. R, Aditya 2015, 'Penggunaan Web Crawler untuk menghimpun Tweet dengan Metode Pre-Processing Text Mining', *Jurnal infotel*, vol. 7, no. 2.
- [3] N, Azam, Abulaish, Jahiruddin, Haldar, M 2015, 'Twitter Data Mining for Events Classification and Analysis', *Proceedings of the 2nd International Conference on Soft Computing and Machine Learning Intelligence (ISCM I'15)*, vol. 10, no.3, hh. 23-24.
- [4] Negara, Edi Surya, Andryani, Ria 2016, 'Analisis Data Twitter: Ekstrasi dan Analisis Data Geospasial', *Jurnal INKOM*, vol. 10, no. 1, hh. 27-36.
- [5] Firstmedia, 2018, *Sejarah First Media*, Tentang Firstmedia, dilihat 3 Desember 2019, <www.firstmedia.com>.
- [6] Praditya, Diaz 2018, *Provider Internet Kabel di Indonesia*, Artikel Tech in Asia, dilihat 5 Desember 2019, <<https://id.techinasia.com/inilah-4-provider-internet-broadband-di-indonesia>>.
- [7] Tim DetikInet, *Layanan Alami Gangguan, Ini Penjelasan First Media*, Artikel Detikinet Telecommunication, dilihat 3 Desember 2019, <https://inet.detik.com/telecommunication/d-4789081/layanan-alami-gangguan-ini-penjelasan-first-media> (accessed on December 3,2019)
- [8] Regina, Luttrell 2015, 'Social Media : How to engage, share and connect', *London : Rowman & Littlefield*, vol. 35 no. 2.
- [9] Rama, Maskeri, Girish, & Kak, Avinash 2013, *Software – Practice and Experience*, Wiley Online Library, New Jersey.
- [10] Kumar, S, Morstatter, F, & Liu, H 2014, 'Twitter Data Analytics', *Journal of Information Technology*, vol.2 no. 11, hh.19-20.
- [11] J.E, Sembodo, E.B, Setiawan, , Z.K.A, Baizal 2016, 'Data Crawling Otomatis pada Twitter', Ind. Symposium on Computing, hh. 11-16.
- [12] B. G., Smith, R.L, Men, &, R, Al-Sinan 2015, 'Tweeing taksim communication power and social media advocacy in the taksim square protest', *Journal of computers in Human Behavior*, no. 50, hh. 499-507.
- [13] Ugheoke, T. O. 2014, *Detecting the Gender of a Tweet Sender*, Department of Computer Science, University of Regina, Canada.
- [14] W, Fan, M.D., Gordon 2014 'The Power of social media analytics', *Communication of ACM*, vol. 57, no. 6, hh.74-81.