

Pelatihan Pengambilan dan Visualisasi Data Twitter Bagi Siswa SMA Intensif Taruna Pembangunan Surabaya

Alfisyahrina Hapsery¹, Artanti Indrasetianingsih², Nanda Mei Sitasari³, Faldianus Karno⁴

^{1, 2, 3, 4} Universitas PGRI Adi Buana Surabaya

*alfisyahrina@unipasby.ac.id

DOI: <https://doi.org/10.31331/manggali.v3i1.2451>

Info Articles

Sejarah Artikel:

Disubmit : November 2022

Direvisi : Desember 2022

Disetujui : Januari 2023

Keywords:

Twitter data, Visualisasi Data, Prapocessing, wordcloud. Data Science.

Abstrak

Pengabdian yang dilakukan dosen Program Studi Statistika Universitas PGRI Adi Buana Surabaya pada semester Gasal 2022/2023 dilaksanakan di SMA Intensif Taruna Pembangunan Surabaya. Kegiatan pengabdian masyarakat merupakan salah satu implementasi dari visi dan misi Program Studi Statistika Universitas PGRI Adi Buana Surabaya yaitu unggul dalam menghasilkan analis data dan saintis data yang berkarakter Peduli, Amanah, Gigih, dan Inovatif. Tujuan kedua untuk mengenalkan proses pengambilan data twitter atau *text mining*. *Data science* juga dikenal dengan nama *data mining* yang diartikan sebuah pengolahan data yang secara volume ataupun struktur data yang tidak seragam. Dasar ilmu data saintis harus mengenal dan menguasai bahasa pemrograman seperti SQL, Python, atau R Studio melalui data *twitter*. Data tersebut dilakukan *prapocessing* dengan menghilangkan awalan, akhiran serta kombinasi awalan dan akhiran. Data tersebut selanjutnya dilakukan statistika deskriptif seperti diagram batang dan *wordcloud*. Kegiatan pelatihan telah terlaksana dengan baik dan lancar. Hasil akhir dari pelatihan ini didapatkan bahwa adanya peningkatan pengetahuan peserta terhadap penggunaan dan proses pengambilan data di akun media sosial *twitter* dengan R studio dan API. Keberhasilan dari kegiatan pelatihan ini dapat dilihat dari hasil simulasi peserta dan pemaparan interpretasi visualisasi data.

Abstract

The service by the lecturers of the Statistics Study Program at PGRI Adi Buana University, Surabaya in the odd semester of 2022/2023 was carried out at the Taruna Pembangunan Intensive High School in Surabaya. Community service activities are one of the implementations of the vision and mission of the Statistics Study Program at PGRI Adi Buana University to producing data analysts and data scientists with Caring, Trustworthy, Persistent, and Innovative characteristics. In addition, the second goal is to

introduce the Twitter data collection process or text mining. Data science is also known as data mining which means data processing that is not uniform in volume or data structure. To become a data scientist, you must know and master programming languages such as SQL, Python, or R Studio. The application of text mining can be done on the Twitter application. Twitter data is preprocessed by removing prefixes, suffixes and combinations of prefixes and suffixes. The data is then carried out with descriptive statistics such as bar charts and word cloud. All stages of the training activities have been carried out well and smoothly. The final result of this training was that there was an increase in participants' knowledge of the use and process of data collection on Twitter social media with R studio and API. The success of this training activity can be seen from the simulation results of the participants and the explanation of the data visualization interpretation.

✉ Alamat Korespondensi:

E-mail: alfisyahrina@unipasby.ac.id

p-ISSN : 2715-5757

e-ISSN : 2798-4435

PENDAHULUAN

Pengabdian Masyarakat adalah suatu kegiatan yang dilakukan dosen dalam memenuhi Tri Dharma Perguruan Tinggi yang diatur pada UUD 12 Tahun 2012. Pengabdian Masyarakat juga merupakan kewajiban yang harus dipenuhi oleh dosen untuk memenuhi salah satu beban kinerja dosen setiap semester. Bentuk pengabdian masyarakat di Prodi Statistika yaitu pelatihan olah data dengan menggunakan software tertentu sesuai kebutuhan peserta pengabdian. Pengabdian yang dilakukan dosen Program Studi Statistika Universitas PGRI Adi Buana Surabaya pada semester Gasal 2022/2023 dilaksanakan di SMA Intensif Taruna Pembangunan Surabaya.

Kegiatan pengabdian masyarakat salah satu implementasi dari visi dan misi Program Studi Statistika Universitas PGRI Adi Buana Surabaya yaitu unggul dalam menghasilkan analisis data dan saintis data yang berkarakter Peduli, Amanah, Gigih, dan Inovatif. Pentingnya data dan *data science* pada era saat ini sehingga diperlukan keterampilan analisis statistika (Hadijati et al., 2021). Hal tersebut disebabkan oleh lonjakan *big data* yang timbul sehingga data yang terkumpul berjumlah besar dan bervariasi (Siregar et al., 2022). *Data science* juga dikenal dengan nama *data mining* yang diartikan sebuah pengolahan data yang secara volume ataupun struktur data yang tidak seragam. Menjadi seorang data saintis harus mengenal dan menguasai bahasa pemrograman seperti SQL, Python, atau R Studio (Hapsery, et al, 2022).

Salah satu implementasi dari *data science* adalah *text mining*. *Text mining* adalah penggalian data untuk menyelesaikan masalah kebutuhan informasi dengan menerapkan teknik *data mining*, *machine learning*, *natural language processing*, pencarian informasi, dan manajemen pengetahuan. Metode ini digunakan untuk mengekstraksi informasi dari sumber data melalui identifikasi dan eksplorasi pola yang menarik (Berry & Kogan, 2010). Penerapan *text mining* dapat dilakukan pada aplikasi *twitter*. *Twitter* adalah sebuah situs jejaring sosial yang sedang berkembang pesat saat ini karena pengguna dapat berinteraksi dengan pengguna lainnya dari komputer ataupun perangkat mobile mereka dari manapun dan kapanpun.

Data yang telah diperoleh dari *twitter* harus melalui proses *prapocessing* data teks agar data dapat mudah untuk dianalisis. *Prapocessing* teks meliputi mengubah semua karakter teks menjadi huruf kecil serta menghilangkan tanda baca dan angka serta mendapatkan kata dasar dengan cara menghilangkan awalan, akhiran, sisipan, dan *confixes* (kombinasi

dari awalan dan akhiran) (Mujilawati, 2016). Setelah data melalui *praprocessing* data teks dapat di analisis menggunakan statistika deskriptif, diagram dan *word cloud*. *Word cloud* adalah representasi grafis dari sebuah dokumen yang dilakukan dengan *plotting* kata-kata yang sering muncul pada sebuah dokumen pada ruang dua dimensi.

Berdasarkan uraian diatas, para dosen dari Prodi Statistika Fakultas Sains dan Teknologi UNIPA Surabaya mengadakan program pengabdian masyarakat dengan memberikan “Pengenalan Eksplorasi dan Analisis Data *Twitter* bagi Siswa SMA Intensif Taruna Pembangunan Surabaya” dengan submateri “Pelatihan Pengambilan dan Visualisasi Data *Twitter* bagi Siswa SMA Intensif Taruna Pembangunan Surabaya”. Kegiatan pengabdian ini merupakan salah satu bentuk kegiatan Tri Dharma Perguruan Tinggi yang harus dilakukan oleh dosen, disamping itu juga merupakan sarana bagi dosen untuk memberikan kontribusi langsung kepada masyarakat sesuai dengan kompetensi keilmuan yang dimiliki.

METODE

Observasi dan sasaran pelatihan ini adalah siswa kelas XII di SMA Intensif Taruna Pembangunan Surabaya. Lokasi Pengabdian di Laboratorium Prodi Statistika UNIPA Surabaya. Lama pelatihan ini dilaksanakan selama satu bulan, dimulai sejak tanggal 17 Oktober 2022 sampai dengan 18 November 2022 secara luring. Teknik pelaksanaan pengabdian bagi program studi statistika yaitu dengan memberikan pelatihan pengambilan data twitter sesuai dengan topik terkini berdasarkan fenomena yang terjadi di bidang sosial.



Gambar 1. Dokumentasi Pelatihan Pengambilan Data *Twitter*

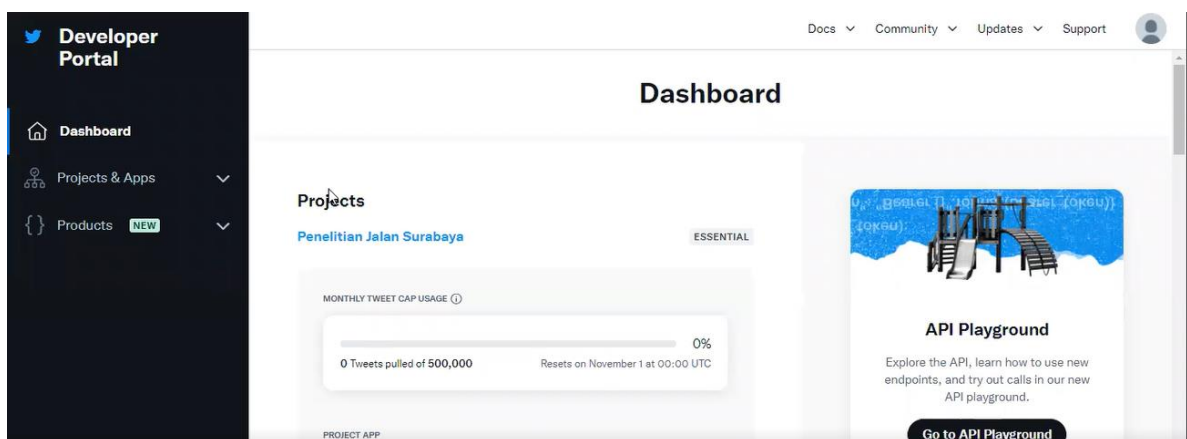
Data twitter yang diperoleh kemudian dilakukan Visualisasi dan proses pengolahan data text yang sering disebut dengan text mining hingga interpretasi hasil visualisasi data text yaitu *wordcloud*. Prosedur pelatihan pengambilan data *twitter* ini dilakukan yaitu diawali dengan metode ceramah atau penjelasan aplikasi yang digunakan, kemudian *workshop* pengambilan data dan visualisasi, prosedur terakhir yaitu simulasi. Berikut adalah dokumentasi dari kegiatan ceramah dan *workshop* Pelaksanaan pelatihan pengambilan dan visualisasi data twitter ini dilakukan secara bertahap. Pertama diawali dengan ceramah, dimana pemateri akan menjelaskan mengenai pengambilan data twitter dengan kata kunci yang terkini yaitu “*Transit Oriented Development*” Kemudian melakukan praproses teks agar data dapat diolah menjadi diagram dan *word cloud*.

Hal kedua yang dilakukan adalah *workshop* dengan data yang diambil dari twitter berdasarkan pemaparan pemateri. Pada tahap ini peserta akan melihat pemateri mempraktekkan pengambilan data *twitter*, proses *cleaning* hingga proses visualisasi. Tahap ketiga yaitu melakukan simulasi, harapa dari tahap ini nantinya peserta mencoba mengambil data *twitter* sesuai dengan tema atau topik yang mereka inginkan sesuai arahan pemateri dan menunjukkan hasil setelah berhasil mencoba. Terakhir dari prosedur pelatihan ini yaitu dengan evaluasi yaitu dengan tanya jawab serta dilakukan pengecekan secara langsung oleh tutor/ pemateri hasil dari simulasi para peserta.

HASIL DAN PEMBAHASAN

HASIL

Siswa XII SMA Intensif Taruna Pembangunan di Surabaya termasuk ke dalam generasi Z dimana saat ini sudah bergerak ke arah revolusi industry 5.0. Selain itu generasi ini harus melek teknologi, khususnya dibidang *Internet of Thing*.

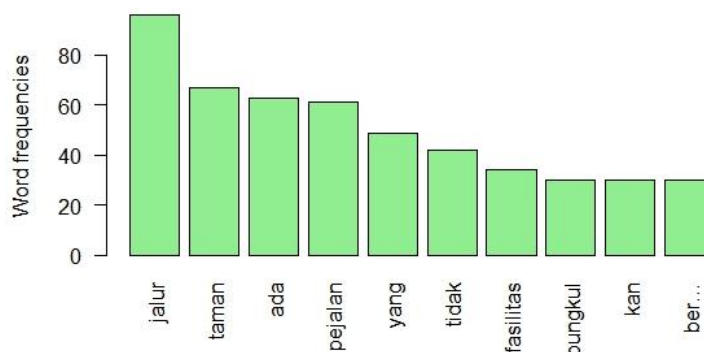


Gambar 1. Application Programming Interface Dashboard

Prodi Statistika membantu memperkenalkan pergerakan IoT dalam bentuk pelatihan pengambilan data di Twitter. Aplikasi yang digunakan dalam proses pengambilan data *twitter* disebut API (*Appllication Programming Interface*) seperti pada Gambar 2 diatas. Terdapat empat hal penting dalam proses pengambilan data twitter yaitu, *consumer keys*, *consumer secret*, *access key*, *access secret*. Tujuannya diadakan pelatihan pengambilan data *twitter* untuk memperkenalkan para siswa generasi Z bahwa pada jaman revolusi industry 5.0 ada beragam jenis data, dan cara memperolehnya.

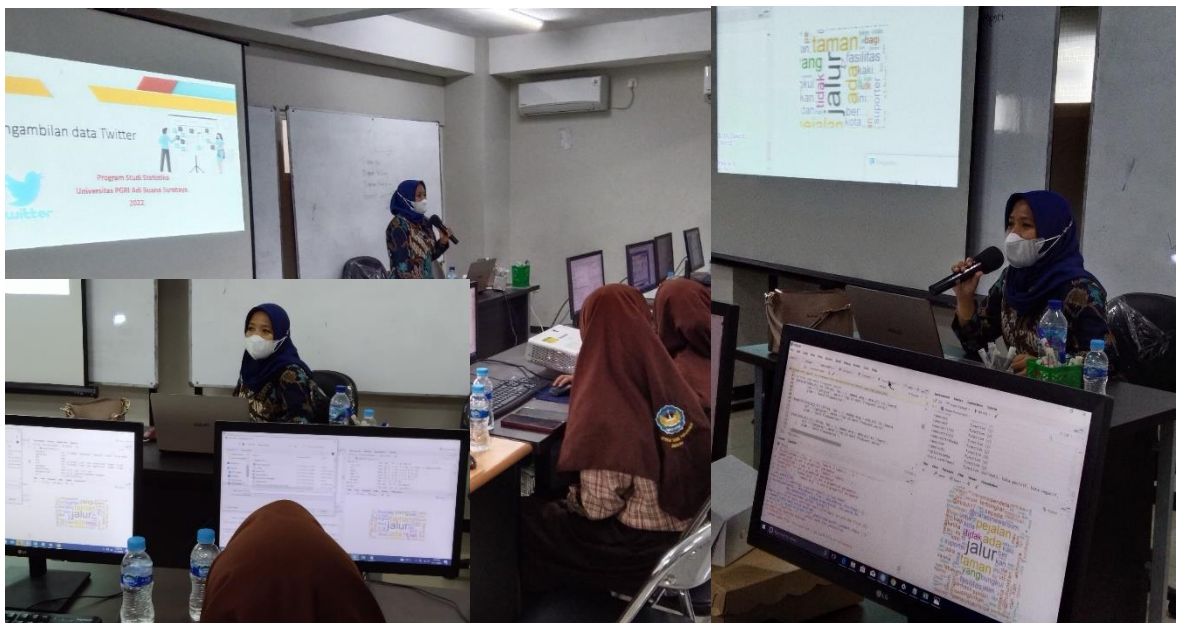
Salah satunya dengan mengambil data melalui persepsi atau pendapat masyarakat di sosial media *twitter*. Sosial media *twitter* merupakan wadah bagi masyarakat untuk bebas mengeluarkan segala persepsi atau pendapat. Akun sosial media twitter akan ramai apabila dalam dunia sosial terjadi beberapa kasus baik hal baik maupun hal buruk yang sedang terjadi detik itu. Misalnya pada bulan November 2022 terjadi kasus yang memakan korban dari tragedi kanjuruhan di Kota Malang. Namun, untuk beberapa hal diluar dari kejadian yang sedang hangat diperbincangkan pada akun twitter juga bisa dilakukan pengambilan data berdasarkan tema atau topik tertentu, seperti “*Transit oriented Development*”.

Transit oriented Development (TOD) merupakan topik yang menceritakan tentang bagaimana tata letak dari suatu taman kota, mulai dari jalur pejalan kaki, ketersediaan tempat duduk, hingga tata letak halte. Data dari akun twitter disebut dengan *text mining*, yang dapat di transformasi menjadi angka melalui pembobot dari nilai frekuensi kata yang muncul. Jumlah kata selain kata penghubung tidak bisa ditentukan, tergantung dari keragaman kata dengan topik TOD. Keragaman data dapat dilihat melalui diagram batang yang telah diurutkan mulai dari kata yang frekuensi muncul lebih banyak dibanding kata lainnya seperti pada Gambar 2.



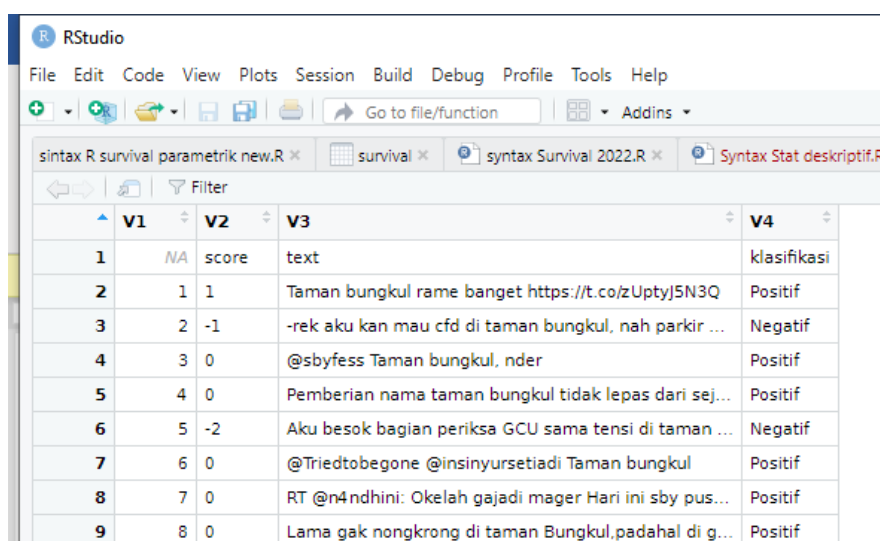
Gambar 2. Sepuluh Kata dengan Frekuensi Tertinggi

Gambar 3 merupakan *wordcloud* dari data TOD. Gambar tersebut menunjukkan bahwa kata yang sering muncul dari status masyarakat dengan tema TOD yaitu jalur, taman, fasilitas, kota, taman bungkul.



Gambar 4. Simulasi Visualisasi Data *Twitter* dengan Tema TOD

Bentuk ini akan mempermudah pembaca untuk segera menemukan kata apa yang paling sering di tulis masyarakat melalui statusnya di akun *twitter*. Penjelasan tentang twitter dan aplikasinya serta worskshop proses running syntax dengan R studio telah dilakukan. Para Siswa SMA ITP selanjutnya melakukan simulasi. Simulasi pertama dilakukan dengan menggunakan tema dan data TOD seperti yang terlihat pada Gambar 4 berikut. Simulasi kedua, para siswa SMA ITP melakukan proses *crawling* data atau pengambilan data dengan tema tertentu sesuai keinginan pada siswa. Pada saat pelatihan, berita yang sedang hangat diperbincangkan adalah kematian sekeluarga di dalam rumah karena kelaparan. Ada juga siswa yang mengambil tema film drama korea. Jumlah data yang tersedia dapat diketahui sebelum dilakukan penyimpanan data. Pada tahap ini, ketika jumlah data tidak sesuai dengan harapan maka kita dapat merubah tema yang akan kita ambil. Jumlah data TOD sebanyak 261 responden atau status tanpa *repost* status milik akun lain.

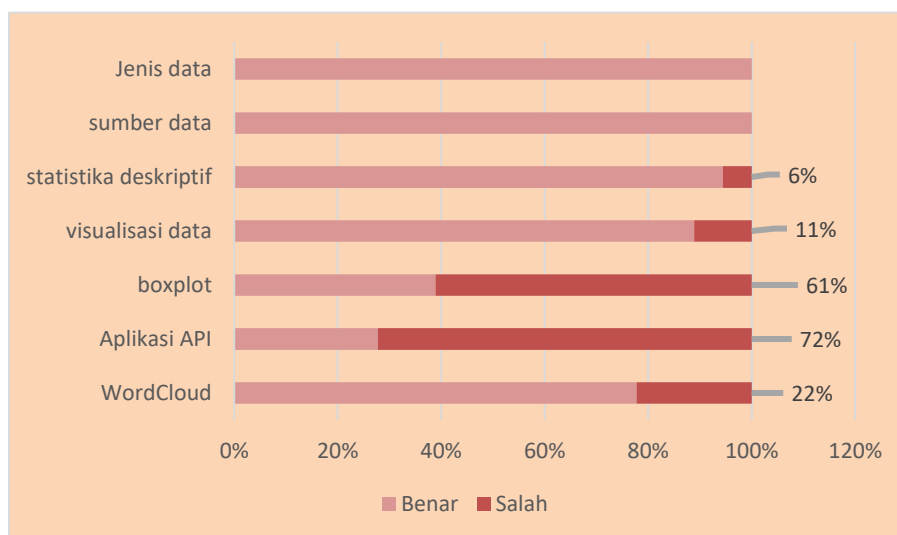


	V1	V2	V3	V4
1	NA	score	text	klasifikasi
2	1	1	Taman bungkul rame banget https://t.co/zUptyJ5N3Q	Positif
3	2	-1	-rek aku kan mau cfd di taman bungkul, nah parkir ...	Negatif
4	3	0	@sbyfess Taman bungkul, nder	Positif
5	4	0	Pemberian nama taman bungkul tidak lepas dari sej...	Positif
6	5	-2	Aku besok bagian periksa GCU sama tensi di taman ...	Negatif
7	6	0	@Triedtobegone @insinyursetiadi Taman bungkul	Positif
8	7	0	RT @n4ndhini: Okelah gajadi mager Hari ini sby pus...	Positif
9	8	0	Lama gak nongkrong di taman Bungkul,padahal di g...	Positif

Gambar 5. Struktur Data Visualisasi Data *Twitter*

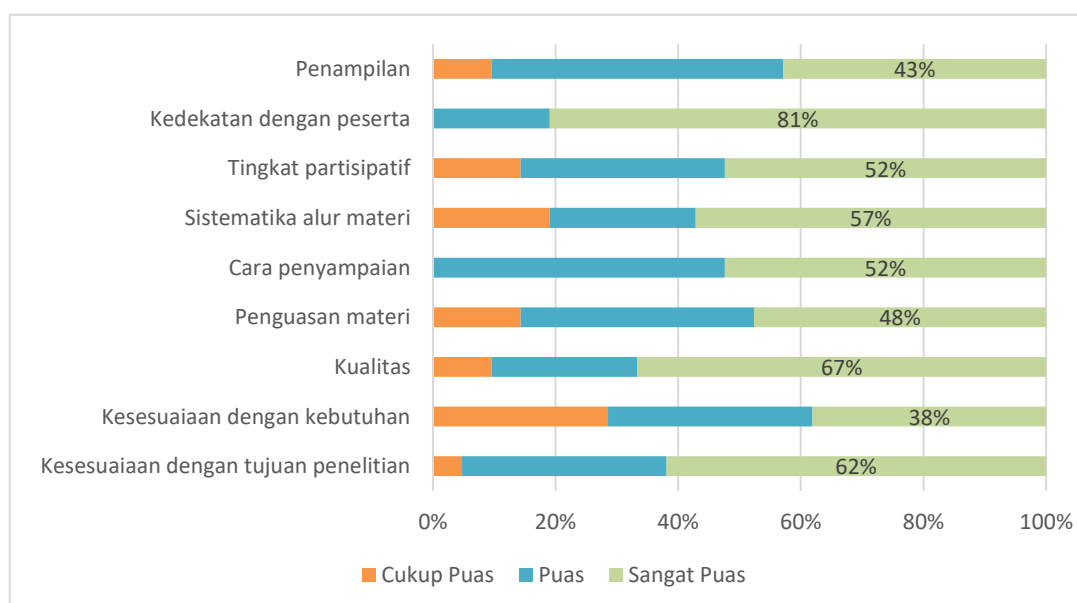
Gambar 5 berikut merupakan bentuk struktur data dilakukan pembobot dan merupakan struktur data untuk visualisasi data. V2 merupakan nilai score yang dihasilkan dari pembobot, V3 merupakan kalimat dari status responden. Sedangkan V4 merupakan hasil klasifikasi dari status responden. Pelatihan ini diharapkan dapat memberi manfaat bagi siswa mengenal data di era digital ini. Setelah pelatihan selesai para peserta diberi soal untuk mengukur pemahaman mereka dari kegiatan pelatihan ini. Selain itu, juga nilai dari test yang dilakukan menjadi tolok ukur pemateri dalam memperbaiki penyampaian, materi, dan bahan pelatihan. Gambar 6 dapat diketahui bahwa siswa SMA kelas XII kurang memahami tentang Aplikasi API, perbedaan

boxplot dengan *wordcloud* untuk data text. Namun, para siswa sudah sangat memahami tentang jenis data, sumber memperoleh data, dan penjelasan mengenai statistika deskriptif. Untuk pengabdian selanjutnya pemateri akan lebih mendalam menjelaskan Aplikasi API untuk pengambilan data dan perbedaan *boxplot* dan *wordcloud*.



Gambar 6. Hasil *Post test* Pelatihan Visualisasi Data Twitter

Gambar 7 berikut adalah nilai kepuasan siswa – siswi SMA ITP Surabaya. Ditinjau dari beberapa aspek penilaian terhadap pemateri dan penilaian terhadap materi yang diberikan, diketahui secara keseluruhan siswa -siswi merasa puas dengan adanya pelatihan ini. Beberapa saran terkait dengan komputer yang sempat mengalami kendala jaringan internet sehingga R Studio tidak bisa digunakan. Mereka juga menilai kegiatan ini sangat sesuai dengan kebutuhan.



Gambar 7. Diagram Batang Perhitungan Kepuasan Pelatihan

SIMPULAN

Pelatihan ini sekaligus ajang promosi bagi prodi untuk mengenalkan dunia data science statistika. Selain itu, tujuan kedua untuk mengenalkan proses pengambilan data *twitter* atau *text mining*. Keseluruhan dari tahapan kegiatan pelatihan telah terlaksana dengan baik dan lancar. Hasil akhir dari pelatihan ini didapatkan bahwa adanya peningkatan pengetahuan peserta terhadap penggunaan dan proses pengambilan data di akun media sosial *twitter* dengan R studio dan API. Keberhasilan dari kegiatan pelatihan ini dapat dilihat dari hasil simulasi peserta dan pemaparan interpretasi visualisasi data.

DAFTAR PUSTAKA

- Berry, M. W., & Kogan, J. 2010. *Text Mining Application and Theory*. United Kingdom: WILEY.
- Hadijati, M., Harsyiah, L., & Fitriyani, N. (2021). Peningkatan Pemahaman Statistika dan Peranannya pada Era Revolusi Industri 4.0 di Madrasah Aliyah Hidayatul Muhsinin Labulia. *Jurnal Karya Pengabdian*, 3(1), 42–46.
- Hapsery, A., Fitriani, F., Hernita, M., Ardiansyah, O. (2022). *Manajemen Data Sekolah SMK Sepuluh Nopember Sidoarjo dengan Menggunakan Excel*, 185–192. <http://e-journal.ivet.ac.id/index.php/manggali>.
- Hapsery, A., Pramesti, W., Khotimah, K., Daimuddin, A. J., (2022). *Visualisasi Data dengan Menggunakan Bahasa Pemrograman R Studio di SMK Informatika Tulangan Sidoarjo*. 41-45. <https://ejournal.stkipgri-sidoarjo.ac.id/index.php/jpadi/issue/view/41>.
- Mujilawati, S. (2016). *Pre-Processing Text Mining Pada Data Twitter*. Seminar Nasional Teknologi Informasi dan Komu- nikasi (SENTIKA), 49-56.
- Siregar, B., Pangruruk, F. A., Siridion, S. T., Immanuel, K. R., Wijaya, J., Supit, V., & Gani, J. I. (2022). Pengenalan Data Science dan Profesi Data Scientist di SMA Pramita Tangerang. *Jurnal Pengabdian Masyarakat Bestari*, 1(2), 87–96. <https://doi.org/10.55927/jpmb.v1i3.620>.