

Implementasi Regresi Linier Untuk Waktu Pengiriman Barang

Lidia Earlene Rendhiva^{1*}, Muhamad Daivany Nur Auilya Saleh², Raditya Koesyan Dipo

Pramukti³, Dwi hartanti⁴

^{1,2,3,4} Teknik Informatika

Universitas Duta Bangsa Surakarta

^{1*}202020976@mhs.udb.ac.id, ²202030149@mhs.udb.ac.id, ³ dwihartanti@udb.ac.id

Abstrak— Belanja online telah menjadi kebiasaan masyarakat dan erat terkait dengan jasa pengiriman barang, yang bertanggung jawab untuk mengirimkan barang yang dibeli dari penjual ke pembeli. Oleh karena itu, pembeli membutuhkan jasa pengiriman yang aman dan cepat untuk memastikan bahwa barang mereka sampai tujuan tepat waktu. Salah satu faktor terpenting dalam bisnis pengiriman adalah kepuasan pelanggan. Namun, beberapa kendala di lapangan memperlambat pengiriman. Oleh karena itu, memprediksi waktu pengiriman dengan menggunakan teknik data mining adalah cara untuk mengatasi masalah ini. Teknik prediksi, algoritma regresi linier, akan digunakan untuk pengolahan data. Dengan menggunakan data tentang kapan barang diambil, saat barang dalam perjalanan, dan sampai di tangan pembeli, mereka dapat membuat perkiraan atau prediksi dan melakukan analisis untuk mencegah keterlambatan pengiriman di masa mendatang. Berdasarkan nilai RMSE (Root Mean Square Error) yang berfungsi untuk menghasilkan nilai level, kesalahan hasil prediksi menggunakan metode ini dan dalam nilai RMSE atau Root Mean Square Error sebesar 0,370%. Dapat disimpulkan bahwa penggunaan algoritma Regresi Linier terbukti akurat dalam memprediksi waktu pengiriman.

Kata Kunci : Data Mining, Regresi Linear, Pengiriman

Abstract— Online shopping has become a habit of society and is closely related to freight forwarders, who are responsible for delivering purchased goods from seller to buyer. Therefore, buyers need a safe and fast shipping service to ensure that their goods arrive at their destination on time. One of the most important factors in the delivery business is customer satisfaction. However, some obstacles in the field slowed down the delivery. Therefore, predicting delivery times using data mining techniques is a way to overcome this problem. The prediction technique, a linear regression algorithm, will be used for data processing. Using data about when the goods were picked up, when they were in transit, and arrived in the hands of buyers, they can make forecasts or predictions and perform analysis to prevent future shipping delays. Based on the RMSE (Root Mean Square Error) value which serves to produce the level value, the error of the prediction results using this method and in the RMSE or Root Mean Square Error value of 0.370%. It can be concluded that the use of Linear Regression algorithms proved to be accurate in predicting delivery times.

Keywords : Data Mining, Linear Regression, Shipping

I. PENDAHULUAN

Dengan belanja online hingga berbagai barang yang dikirim tiba tepat waktu dan pengiriman pribadi dan perusahaan. Setiap konsumen membutuhkan jasa pengiriman yang cepat dan aman untuk memastikan selamat sampai di tempat tujuan. Sudah banyak perusahaan jasa pengiriman yang muncul dan kemudian berkembang untuk bersaing merebut pasar di era modern seperti sekarang ini, dengan tingkat perkembangan ekonomi dan teknologi yang sangat pesat, salah satunya adalah PT EXM yang merupakan salah satu perusahaan yang bergerak di bidang pengiriman barang dengan memiliki kantor cabang yang tersebar di setiap kota di Indonesia. Karena banyaknya barang yang harus dikirim, PT. EXM mengalami kesulitan dalam mencapai waktu

tempuh yang ditentukan hingga barang diterima oleh pelanggan demi memberikan kepuasan pelanggan. Pengiriman yang efisien akan berdampak positif pada pengembangan bisnis. Barang yang menjangkau pelanggan lebih cepat tentunya akan memberikan ruang bagi perusahaan untuk melakukan quality control terhadap barang yang dikirim, sehingga perusahaan dapat berkembang. Sehingga diperlukan metode Data Mining yang dapat melengkapi dan mengoptimalkan kinerja dengan tujuan untuk meningkatkan kepuasan pelanggan. Berdasarkan pembahasan kondisi tersebut, penulis mencoba menawarkan strategi untuk menghadapi keterlambatan pengiriman barang, menggunakan alat aplikasi Rapid Miner dan analisis Data Mining Prediction dengan konsep Linear Regression Algorithm.

Systematic Literature Review (SLR) adalah metode kajian pustaka sistematis yang berfungsi untuk mengidentifikasi, dan menginterpretasikan temuan pada suatu topik penelitian untuk menjawab pertanyaan penelitian yang telah ditentukan. Metodologi survei ini didasarkan pada PICOC (Population, Intervention, Comparison, Outcomes, Context) sebagai identifikasi kebutuhan informasi dari sumber penelitian sebelumnya pada tabel berikut

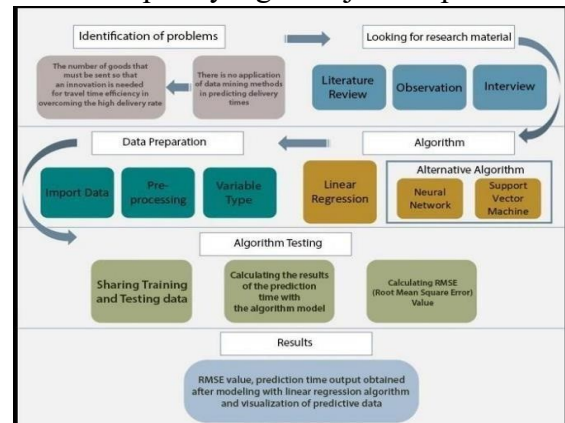
Table 1 - PICOC review

Implementation Of Data Mining Prediction Delivery Time Using Linear Regression Algorithm	
Population	Linear Regression
Intervention	Predicting Freight Forwarding
Comparison	n/a
Outcomes	Predict Delivery of Goods on Time or Not
Context	Private

II. METODOLOGI PENELITIAN

Algoritma yang digunakan dalam penelitian ini adalah Algoritma Regresi Linear. Data pengiriman akan diproses untuk mendapatkan nilai kesalahan dan dapat digunakan sebagai acuan dalam memprediksi waktu pengiriman barang. Jenis penelitian yang dilakukan peneliti adalah penelitian kualitatif. Penelitian kualitatif adalah penelitian yang menggunakan lebih banyak analisis. Proses yang lebih ditekankan dalam jenis penelitian ini adalah landasan teori yang digunakan sebagai pedoman agar fokus penelitian sesuai dengan fakta di lapangan. Selain itu, landasan teori juga memiliki peran untuk memberikan gambaran latar belakang penelitian dan sebagai bahan pembahasan hasil penelitian. Dalam penelitian ini, proses dan makna yang digunakan lebih fokus pada penelitian berdasarkan fakta di lapangan. Analisis data dalam penelitian kualitatif dimaknai sebagai upaya mencari dan menyusun catatan secara sistematis dari pengamatan, wawancara, dan studi pustaka untuk meningkatkan pemahaman peneliti terhadap kasus yang diteliti dan menyajikannya sebagai temuan. Secara umum, penelitian kualitatif memperoleh data dari wawancara dan observasi. Kemudian peneliti kemudian akan menganalisis data yang diperoleh secara detail, dan menghasilkan teori atau konsep baru jika hasil penelitian bertentangan dengan teori dan konsep yang digunakan. Dalam penelitian ini, beberapa tahapan penelitian akan dilakukan. Dalam

penelitian ini, beberapa tahap penelitian akan dilakukan seperti yang ditunjukkan pada Gambar 1:



Gambar. 1. Tahapan Metodologi Aplikasi

Pengumpulan data juga didukung oleh observasi dan wawancara dengan pihak-pihak terkait di lapangan serta pengumpulan data pustaka sebagai referensi dari penelitian sebelumnya.

1. Observasi

Observasi adalah teknik pengumpulan data yang dilakukan dengan cara mencatat dan mengamati PT. EXM secara sistematis. Dengan studi observasi lapangan ini, peneliti memperoleh materi yang cukup akurat dan relevan serta sangat membantu dalam penulisan penelitian ini.

2. Wawancara

Wawancara adalah metode atau teknik yang digunakan untuk memperoleh data dengan melakukan tanya jawab lisan langsung kepada Pemimpin Toko Pementasan. Bentuk informasi yang diperoleh bisa berupa tulisan, atau rekaman audio visual.

3. Tinjauan Pustaka

Tinjauan pustaka: Studi pustaka atau studi pustaka merupakan langkah awal dalam kegiatan pengumpulan data. Hal ini dapat dilakukan dengan membaca jurnal ilmiah, mencatat dan mengolah literatur, buku, dan laporan bahan penelitian sebelumnya untuk memperoleh informasi dan mencari data di internet. Di bawah ini adalah beberapa hasil definisi istilah-istilah penting yang diperoleh dari studi literatur dan terkait penelitian:

3.1. Pengembangan Data

Penambangan Data Data Mining adalah proses untuk memperoleh informasi yang berguna dari database besar yang perlu diekstraksi sehingga

menjadi informasi baru dan dapat membantu dalam pengambilan keputusan.

3.2. Prediksi

Menurut Kamus Besar Bahasa Indonesia, Prediksi adalah hasil dari memprediksi atau memperkirakan nilai masa depan menggunakan data masa lalu. Prediksi menunjukkan apa yang akan terjadi dalam situasi tertentu dan merupakan masukan untuk proses perencanaan dan pengambilan keputusan. Pengertian Prediksi sama dengan ramalan.

3.3. Regresi Linier

Regresi Linier digunakan untuk memperkirakan atau memprediksi hubungan antara dua variabel dalam penelitian kualitatif. Pendekatan untuk menetapkan hubungan antara satu atau lebih variabel dependen (regresi linier sederhana) serta variabel independen (regresi linier berganda). Dengan asumsi bahwa hubungan antara variabel-variabel ini dapat didekati dengan persamaan garis lurus, model yang mendekati hubungan antara variabel dalam data disebut sebagai stabilisasi regresi linier.

3.4. Dataset

Dataset atau Kumpulan Data adalah kumpulan data yang digunakan sebagai tujuan dari beberapa pembelajaran untuk mesin tertentu. Ada 2 kategori himpunan data: publik dan pribadi. Himpunan data memiliki atribut yang berfungsi sebagai faktor atau parameter yang menyebabkan kelas, label, dan target terjadi. Tren penelitian data mining saat ini adalah menguji metode yang dikembangkan oleh peneliti dengan dataset publik, sehingga penelitian dapat dibandingkan, diulang dan diverifikasi.

3.5. CRISP-DM

CRISP-DM (Cross Industry Standard Process for Data Mining) Merupakan standarisasi pengolahan data mining yang telah dikembangkan dimana data yang ada akan melewati setiap fase terstruktur dan terdefinisi secara jelas dan efisien.

3.6. Rapidminer

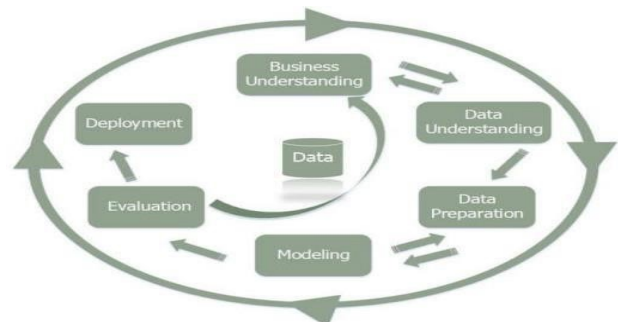
Adalah aplikasi Data mining berbasis sumber terbuka terkemuka dan terkenal. Kegunaan aplikasi

ini adalah untuk analisis data dan sebagai mesin penambangan data seperti untuk pemuatan data, transformasi data, pemodelan data, dan metode visualisasi data.

III. HASIL DAN PEMBAHASAN

Berdasarkan tahapan penerapan metodologi, peneliti menerapkan desain DM Crisp sebagai model uji, dan Algoritma Regresi Linier sebagai metode data mining dalam penelitian ini. Dalam penelitian ini, desain pengujian yang dilakukan dengan menggunakan CrossIndustry Standard Process for Data Mining atau CRISP-DM merupakan salah satu model proses datamining (datamining framework).

Berikut ini adalah gambar skema pengujian pada CRISPDM:



Gambar 2. CRISP DM

Skema Uji Dalam penelitian ini menggunakan model uji untuk mendapatkan hasil penelitian yang baik dan maksimal. Sebelum melakukan pengolahan data dan pengujian algoritma Regresi Linear, langkah pertama yang harus dilakukan adalah menerapkan model uji data mining. Data yang diproses oleh model uji CRISP-DM akan melalui beberapa fase di dalamnya. Terdapat 6 tahapan dalam CRISP-DM, yang dijelaskan secara rinci sebagai berikut:

1. Tahapan Pengertian Bisnis Jasa pengiriman barang baik dalam lingkup nasional maupun internasional saat ini dituntut untuk memiliki keunggulan dalam menjadi kompetitif, yaitu dengan meningkatkan kualitas layanannya. Agar pelayanan yang baik tercapai, maka dipengaruhi oleh kecepatan pengiriman barang yang tepat waktu dan aman ketika sampai di penerima. Dengan banyaknya paket, waktu pengiriman berbeda, sehingga ada yang datang tepat waktu dan ada yang tidak. Pengiriman barang yang tidak

tepat waktu bisa menjadi catatan penting bagi perusahaan pelayaran.

2. Tahapan Pemahaman Data Semua data yang telah dikumpulkan adalah 1000 data. Data yang digunakan dalam penelitian ini adalah data pengiriman barang di PT. EXM dalam rentang waktu 1 hari. Data ini dikumpulkan dari sistem informasi yang ada di divisi Management Information System (MIS). Tidak semua atribut data pengiriman digunakan, peneliti hanya menggunakan 3 atribut yang terdiri dari Pickup Time, Claim Delivery Time, dan Delivered Time. Atribut AWB tidak digunakan karena atribut ini tidak diperlukan dalam data penelitian untuk memprediksi waktu pengiriman, sehingga hanya membutuhkan 3 atribut. Berikut adalah rincian semua atribut:

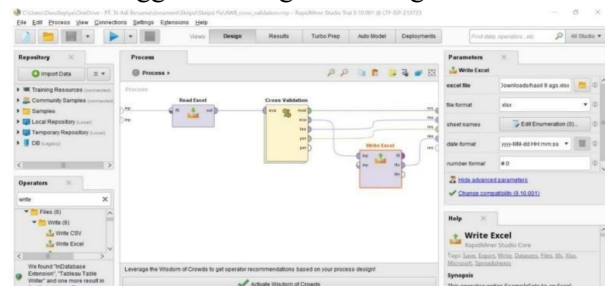
- Waktu Penjemputan atau Goods Pickup Time adalah waktu layanan pengambilan barang sehingga menggunakan tipe data tanggal.
- Waktu Klaim Pengiriman atau Time of Claim Delivery adalah waktu ketika kurir mulai mendistribusikan paket barang dan memiliki tipe data tanggal waktu.
- Waktu Pengiriman atau waktu ketika barang telah tiba di tempat tujuan dan diterima oleh penerima menggunakan tipe data tanggal, waktu.

3. Tahapan Persiapan Data Pada tahap ini, data disusun terlebih dahulu dengan menggunakan teknik pre-processing, yaitu data mentah diubah menjadi bentuk yang lebih mudah dipahami. Dengan menggunakan teknik ini data dibersihkan, atribut yang tidak perlu digunakan dalam proses analisis, seperti atribut AWB, akan dihapus. Berikut langkah-langkah dalam melakukan pre- processing:



Gambar 3. Memilih Atribut dan Jenis Variabel

4. Tahapan Pemodelan Pada tahap pemodelan, adalah membuat model prediksi, yaitu memprediksi waktu pengiriman barang. Pada tahap ini, Anda dapat menggunakan read excel untuk mengimpor kumpulan data, validasi silang untuk berbagi data pelatihan dan pengujian, dan excel writer untuk mengeksport hasil pemrosesan data dalam proses validasi silang. Berikut ini adalah model proses yang dibentuk pada Rapid Miner menggunakan algoritma Regresi Linear:



Gambar 4. Tampilan Model Prediksi

Regresi linier umumnya akan membuat prediksi berdasarkan nilai yang sudah ada sebelumnya.

Attribute	Coefficient	Std Error	Sig. Coefficient	Inference	t Stat	P Value	Code
pickup_time	0.135	0.027	0.000	0.000	4.969	0.000	---
claim_delivery_time	0.744	0.020	0.000	0.000	36.666	0.000	---
delivered_time	0.003	0.000	0.000	0.000	6.112	0.000	---

Gambar 5. Hasil Regresi Linier

5. Tahapan Evaluasi Setelah pemodelan, langkah selanjutnya adalah mengevaluasi proses cross validation dengan menerapkan algoritma regresi linier, cara kerjanya adalah mevaluasi kinerja algoritma dimana data dipisahkan menjadi dua subset, yaitu data proses pembelajaran (training) dan validasi data (testing). Berikut ini adalah hasil evaluasi model yang telah dibuat dengan algoritma ini.



Gambar 6. Tahapan Evaluasi Validasi Silang

Operator performance digunakan untuk mengevaluasi kinerja model yang menyediakan daftar nilai kriteria kinerja secara otomatis sesuai dengan tugas yang diberikan. Dari vektor pada Gambar 5, nilai kesalahan yang diperoleh adalah 0,370. Pengukuran kinerja dilakukan dengan menghitung kesalahan rata-rata yang terjadi melalui jumlah Root Mean Square Error (RMSE). Semakin kecil nilai masing-masing parameter kinerja ini, semakin dekat nilai prediksi dengan nilai actual.



Gambar 7. Hasil Nilai Nilai RMSE

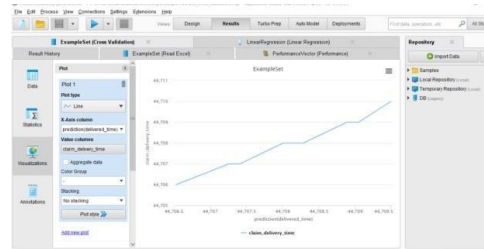
Berikut ini adalah ekspor hasil prediksi waktu yang telah dikonversi ke dalam bentuk excel. Pada data di bawah ini, hasil prediksi waktu pengiriman tidak jauh berbeda dengan hasil data aktual pada waktu pengiriman.

	A	B	C	D
	pickup_time	claim_delivery_time	delivered_time	prediction(delivered_time)
1	25/05/2022 18:50	26/05/2022 13:15	26/05/2022 14:38	26/05/2022 15:09
2	25/05/2022 13:48	26/05/2022 16:05	26/05/2022 17:32	26/05/2022 17:52
3	25/05/2022 13:43	26/05/2022 11:24	26/05/2022 12:41	26/05/2022 13:13
4	25/05/2022 14:46	26/05/2022 16:36	26/05/2022 17:34	26/05/2022 18:34
5	25/05/2022 13:34	26/05/2022 16:29	26/05/2022 18:42	26/05/2022 18:55
6	25/05/2022 19:11	26/05/2022 10:21	26/05/2022 12:50	26/05/2022 12:17
7	25/05/2022 14:35	26/05/2022 12:01	26/05/2022 12:52	26/05/2022 13:50
8	25/05/2022 19:09	26/05/2022 15:19	26/05/2022 15:37	26/05/2022 17:13
9	25/05/2022 10:09	26/05/2022 13:24	26/05/2022 13:32	26/05/2022 15:07
10	25/05/2022 12:31	26/05/2022 14:48	26/05/2022 15:58	26/05/2022 16:33
11	25/05/2022 10:48	26/05/2022 17:22	26/05/2022 17:49	26/05/2022 19:12
12	25/05/2022 16:23	26/05/2022 15:53	26/05/2022 16:30	26/05/2022 17:43
13	25/05/2022 16:56	26/05/2022 10:30	26/05/2022 11:52	26/05/2022 12:13
14	25/05/2022 14:34	26/05/2022 14:45	26/05/2022 15:52	26/05/2022 16:34
15	25/05/2022 20:08	26/05/2022 12:59	26/05/2022 14:34	26/05/2022 14:55
16	25/05/2022 11:47	26/05/2022 11:17	26/05/2022 13:36	26/05/2022 13:03
17	25/05/2022 19:28	26/05/2022 10:45	26/05/2022 20:55	26/05/2022 12:42
18	25/05/2022 11:12	26/05/2022 15:52	26/05/2022 16:54	26/05/2022 17:36

Gambar 8. Waktu Prediksi Hasil

6. Tahap Penyebaran Tahap Deployment dilakukan setelah tahap evaluasi. Penilaian terperinci dari hasil model dilakukan dengan implementasi seluruh model yang telah dibangun. Selain itu, dilakukan penyesuaian model sehingga dapat menghasilkan hasil yang sesuai

dengan target awal penelitian ini. Agar data prediksi waktu pengiriman berdasarkan data pengujian dan pengujian yang dimodelkan mudah dipahami, maka perlu dilakukan visualisasi data tersebut ke dalam diagram dashboard atau tampilan grafik.



Gambar 9. Menampilkan Visualisasi Data

IV. KESIMPULAN

Berdasarkan hasil penelitian dari tersebut, dapat disimpulkan bahwa: Data olahan adalah data yang diperoleh dalam 1 hari pengiriman dengan total 1000. Data yang diproses menghasilkan nilai koefisien untuk waktu pengambilan sebesar 0,135 dan waktu pengiriman klaim sebesar 0,744. Faktor yang mempengaruhi hasil prediksi adalah atribut waktu pengambilan dan waktu pengiriman klaim, dengan menjadikan waktu pengiriman sebagai label. Hasil yang terbentuk pada proses prediksi waktu yang disampaikan dengan Algoritma Regresi Linier menghasilkan RMSE sebesar 0,370, maka dapat disimpulkan bahwa Semakin kecil nilai RMSE dari masing-masing parameter kinerja tersebut menunjukkan semakin dekat nilai prediksi dengan nilai asli.

REFERENSI

- [1] Ahmad, R., & Alsmadi, I. (2021). Pendekatan pembelajaran mesin untuk keamanan IoT: Tinjauan literatur yang sistematis. Internet Hal, 14, 100365.
- [2] PLOTNIKOVA V, DUMAS M, and MILANI F. Adaptations of data mining methodologies: a systematic literature review, PeerJ Computer Science, 2020; 6: e267, <https://doi.org/10.7717/peerj-cs.267>.
- [4] SCHÄFER F, ZEISELMAIR C, and BECKER J. Synthesizing CRISP-DM and Quality Management: A Data Mining Approach for Production Processes 2020
- [5] IEEE International Conference on Technology Management, Operations and Decisions (ICTMOD), 2018: 190-195, <https://doi.org/10.1109/ITMC.2018.8691266>.
- [6] MARTÍNEZ-PLUMED F, OCHANDO L. C, FERRI C, et al. CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. IEEE Transactions on Knowledge and

- [7] Data Engineering, 2021, 33(8): 3048- 3061, <https://doi.org/10.1109/tkde.2019.2962680>.
- [8] WISEMAN V, THABRANY H, ASANTE A. et al. An evaluation of health systems equity in Indonesia: study protocol. *International Journal for Equity in Health*, 2018, 17(1): 138. <https://doi.org/10.1186/s12939-018-0822-0>
- [9] ELTAHIR O A B. The Effect of Information Technology on The Cooperative Insurance Industry Case Study: Shiekan Insurance and Reinsurance Company – Sudan perception. *Bulletin of Electrical Engineering and Informatics*, 10(4), 2231- 2236.
- [10] Hafizah, Tugiono, & Maya, W. R. (2019). Penerapan Data Mining Dalam Memprediksi Jumlah Penumpang Pada CV . Surya Mandiri Sukses Dengan Menggunakan Metode Regresi Linier. *Jurnal Teknologi Informasi Dan Sistem Komputer TGD*, 2(1), 54–61.
- [11] Bengnga, A., & Ishak, R. (2018). Prediksi Jumlah Mahasiswa Registrasi Per Semester Menggunakan Linier Regresi Pada Universitas Ichsan Gorontalo. *ILKOM Jurnal Ilmiah*, 10(2), 136–143. <https://doi.org/10.33096/ilkom.v10i2.274.136-143>
- [12] Artin, J., Valizadeh, A., Ahmadi, M., Kumar, S. A., & Sharifi, A. (2021). Presentation of a novel method for prediction of traffic with climate condition based on ensemble learning of neural architecture search (NAS) and linear regression. *Complexity*, 2021.
- [13] (Empirical Study). *International Journal of Economics, Business and Accounting Research*, 2020, 4(1): 27 – 37, <https://jurnal.stieaas.ac.id/index.php/IJEBAR>
- [14] HIWASE V A, and AGRAWA A J. Review on Application of Data Mining in Life Insurance. *International Journal of Engineering & Technology*, 2018, 7(45):159- 162, <http://dx.doi.org/10.14419/ijet.v7i4.5.20035>.
- [15] Hertina, H., Nurwahid, M., Haswir, H., Sayuti, H., Darwis, A., Rahman, M., ... & Hamzah, M. L. (2021). Data mining applied about polygamy using sentiment analysis on Twitters in Indonesian