

BAB II

TINJAUAN PUSTAKA

2.1 SEVENTEEN

SEVENTEEN adalah boy grup asal Korea Selatan dibawah naungan agensi Pledis Entertainment. SEVENTEEN memulai debutnya pada bulan Mei 2015 yang terdiri dari 13 anggota: S. Coups (Choi Seung Cheol), Jeonghan (Yoon Jeonghan), Joshua (Joshua Hong), Jun (Wen Junhui), Hoshi (Kwon Soon Young), Wonwoo (Jeon Won Woo), Woozi (Lee Ji Hoon), DK (Lee Seok Min), Mingyu (Kim Min Gyu), The8 (Xu Minghao), Seungkwan (Boo Seung Kwan), Vernon (Hansol Vernon Chwe), dan Dino (Lee Chan) (Koivukoski, 2023).



Gambar 2. 1 Grup SEVENTEEN

Situs akumulasi penjualan musik, Circle Chart, melaporkan bahwa SEVENTEEN telah menjual lebih dari satu juta kopi album lamanya tahun ini. Menurut laporan, album Seventeen yang dirilis sebelum tahun 2024 telah

terjual lebih dari 1,08 juta kopi dari 1 Januari hingga 6 Juli. SEVENTEEN adalah artis K-pop pertama di tahun ini yang berhasil melampaui penjualan sejuta kopi album lawas, menurut agensi SEVENTEEN, Pledis Entertainment. Tahun lalu, Seventeen menjual sekitar 4,2 juta kopi album lama. Seventeen mencatat lebih dari 5,3 juta penjualan album pada paruh pertama tahun ini. Tahun lalu, penjualan album kumulatif Seventeen mencapai 16 juta kopi.

2.2 Prediksi

Prediksi adalah suatu proses memperkirakan secara sistematis tentang sesuatu yang paling mungkin terjadi di masa depan berdasarkan informasi masa lalu dan sekarang yang dimiliki, agar kesalahannya (selisih antara sesuatu yang terjadi dengan hasil perkiraan) dapat diperkecil. Prediksi tidak harus memberikan jawaban secara pasti kejadian yang akan terjadi, melainkan berusaha untuk mencari jawaban sedekat mungkin yang akan terjadi (Andreas, 2019).

Prediksi menunjukkan apa yang akan terjadi pada suatu keadaan tertentu dan merupakan input bagi proses perencanaan dan pengambilan keputusan. (Kafil, 2019). Penerapan prediksi secara luas dalam berbagai disiplin ilmu memungkinkan perencanaan yang lebih efektif dan pengambilan keputusan yang lebih rasional dalam menghadapi ketidakpastian masa depan.

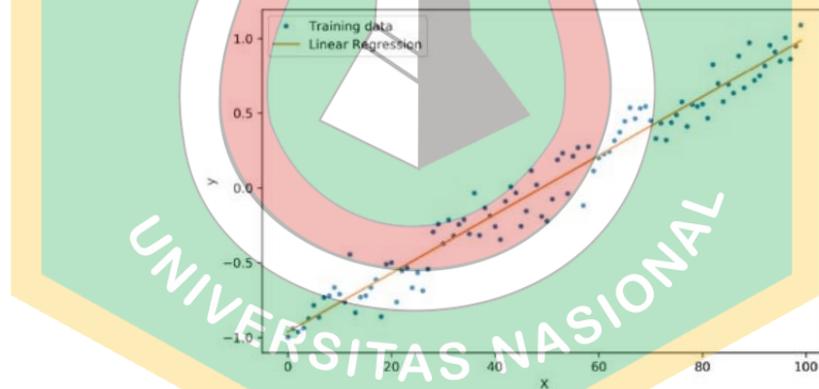
2.3 K-Nearest Neighbor

K-Nearest Neighbour (KNN) adalah salah satu algoritma dalam data mining yang digunakan untuk mengklasifikasikan objek berdasarkan data pelatihan dengan jarak terdekat dari objek tersebut (Ema Utami et al., 2019). Menurut (Siti Nur Asiyah & Kartika Fithriasari, 2016) algoritma KNN dikenal karena efisiensinya yang tinggi dan dalam banyak kasus mampu

memberikan akurasi yang baik dalam proses prediksi. Prinsip dasar dari KNN adalah menentukan kategori atau nilai suatu data baru dengan mengacu pada kemiripan atribut antara data tersebut dengan sekumpulan data yang telah dilabeli sebelumnya. Meskipun pada awalnya dikembangkan sebagai metode klasifikasi, dalam beberapa tahun terakhir algoritma KNN juga telah diaplikasikan secara luas untuk keperluan prediksi karena kesederhanaannya dan kemampuannya dalam menangkap pola dalam data.

2.4 Regresi Linier

Regresi linier adalah salah satu metode yang paling sederhana dan paling mudah digunakan dalam prediksi (Kuriakose et al., 2020). Sebuah model regresi linier awalnya belajar menggunakan data latih di mana ia mempelajari hubungan linier antara input dan output data dan setelah dilatih dapat digunakan pada data yang tidak terlihat. Regresi linier membantu menemukan hubungan linier antara variable independen (*predictor*) dan variable dependen (*target*) seperti yang ditunjukkan pada gambar.



Gambar 2. 2 Peran regresi linier

2.5 Machine Learning

Machine learning adalah aplikasi atau bagian dari *Artificial Intelligence* (AI) yang membuat sistem memiliki kemampuan belajar secara otomatis dan

meningkatkan kemampuannya berdasarkan pengalaman tanpa diprogram secara eksplisit. Fokus *machine learning* terdapat pada pengembangan program komputer yang dapat mengakses data dan belajar dari data tersebut. Dapat diartikan juga bahwa *machine learning* sebagai metode komputasi dengan menggunakan pengalaman untuk meningkatkan kinerja atau membuat prediksi secara akurat (Kusuma, 2020).

2.6 Data Mining

Dalam beberapa tahun terakhir, banyak penelitian telah dilakukan pada *data mining* yang menunjukkan kelebihan dan kekurangan dalam penelitian analisis sentiment. *Data mining* adalah bidang studi dalam *machine learning* dan berfokus pada eksplorasi data analisis data melalui *unsupervised learning*. *Data mining* digunakan untuk menemukan pola dan hubungan dalam data, dengan penekanan pada basis data observasi yang besar (*Data Mining and Statistics*, 2024). (R. Kurniawan & Apriliani, 2020) juga berpendapat *data mining* adalah proses analisis dari berbagai pandangan untuk mendapatkan kesimpulan sehingga menjadi informasi berharga yang dapat meningkatkan keuntungan dengan biaya yang minimal.

2.7 Bahasa Pemrograman Python

Python merupakan bahasa pemrograman tingkat tinggi yang dibuat oleh Guido Van Rossum dan dirilis pada tahun 1991 (M. Riziq Sirfatullah Alfarizi et al., 2023). Python dirancang dengan fokus pada keterbacaan kode, yang menjadikannya mudah dipelajari dan digunakan oleh pemula maupun programmer berpengalaman. Bahasa ini mendukung berbagai paradigma pemrograman, termasuk pemrograman berorientasi objek, prosedural, dan fungsional, serta dengan sintaks yang sederhana dan pustaka seperti pandas, numpy, dan scikit-learn, Python mempermudah pengolahan data, penerapan algoritma, dan evaluasi model secara efisien.

2.8 Visual Studio Code

Menurut (Uci Juwita Agustiani et al., 2021) Visual Studio adalah *integrated development environment (IDE)* yang dikembangkan oleh Microsoft mempermudah *software developer* mengembangkan aplikasi pada platform milik Microsoft. Visual Studio Code adalah *source code editor multiplatform* yang dapat digunakan pada sistem operasi Windows, Linux dan Mac OSX Visual Studio Code juga mendukung banyak bahasa pemrograman. Dengan fitur seperti ekstensi, *debugging*, dan integrasi terminal, Visual Studio Code mempermudah proses pengembangan, terutama dalam menulis dan menguji kode Python untuk analisis data dan pengembangan model *machine learning*.

2.9 Streamlit

Framework Open-Source Streamlit, yang sangat populer dalam komunitas *data science* dan *machine learning*. *Framework* ini memungkinkan pengembang untuk mengubah skrip Python menjadi aplikasi *web* interaktif dengan antarmuka *web*. *Framework* ini sangat cocok untuk penelitian ini karena dimaksudkan untuk membantu membangun aplikasi *data mining* yang melibatkan analisis *data science* dan sistem pembelajaran mesin lainnya (Sarangpure et al., 2023).

Kemudahan membuat aplikasi *web* tanpa memahami teknologi *web* secara mendalam adalah salah satu keunggulan utama Streamlit. Pengembang dapat berkonsentrasi pada logika bisnis dan implementasi model pembelajaran mesin dengan Streamlit, karena *framework* ini secara otomatis menangani elemen antarmuka *web*. Selain itu, fitur-fitur yang sangat bermanfaat untuk aplikasi *data mining* seperti visualisasi data interaktif dan pemetaan interaktif, serta integrasi dengan berbagai pustaka populer dalam ekosistem Python, seperti scikit-learn, pandas, dan NumPy, memungkinkan

pengembang untuk menyajikan hasil analisis data dan prediksi model *machine learning* dalam bentuk yang lebih interaktif dan mudah dipahami bagi pengguna (Syafarina & Zaenuddin, 2023).

2.10 Uji Instrumen Data

Uji Instrumen Data adalah proses pengujian yang dilakukan untuk memastikan bahwa alat pengumpul data (seperti kuisioner) valid dan reliabel, sehingga data yang diperoleh akurat dan dapat dipertanggungjawabkan. Validitas mengacu pada ketepatan alat ukur dalam mengukur apa yang seharusnya diuku, sedangkan realibilitas menunjukkan sejauh mana alat pengukur dapat dipercaya dan konsisten dalam menghasilkan data yang sama dalam waktu yang berbeda. Uji validitas bertujuan untuk melihat ketepatan pengukuran. Sedangkan uji realibilitas bertujuan untuk mengetahui sejauh mana instrument tersebut dapat dipercaya (Ina Marthiani, 2024).

2.11 Uji Asumsi Klasik

Pemenuhan uji asumsi klasik merupakan suatu keharusan dalam metode statistic berdasarkan prinsip analisis regresi linear berganda Ordinary Least Squares (OLS). Pengujian ini dilakukan untuk memastikan bahwa model regresi yang digunakan memiliki estimasi yang akurat, bebas dari bias, serta menunjukkan konsistensi yang baik (Sholihah et al., 2023). Dengan adanya uji asumsi klasik, dapat dipastikan bahwa model regresi yang diterapkan sesuai dengan kaidah yang benar dan dapat menghasilkan hasil analisis yang valid.

2.12 Korelasi Pearson

Korelasi Pearson merupakan korelasi hanya melibatkan (dependen) satu sederhana variabel yang terikat dan satu variabel bebas (independen). Korelasi Pearson menghasilkan koefisien korelasi yang berfungsi untuk

mengukur kekuatan hubungan linier antara dua variabel. Jika hubungan dua variabel tidak linier, maka koefisien korelasi Pearson tersebut tidak mencerminkan kekuatan hubungan dua variabel yang sedang diteliti, meski kedua variabel mempunyai hubungan kuat (Safitri, 2016)

Dalam penelitian ini, analisis korelasi Pearson diterapkan untuk mengevaluasi pengaruh variabel independen seperti Youtube, Concert, Events, dan Fanmeeting terhadap variabel dependen, yaitu Sales. Korelasi ini memberikan pengetahuan lebih tentang variabel mana yang memiliki kontribusi signifikan dalam prediksi penjualan.

2.13 MAE

Mean Absolute Error (MAE) merupakan salah satu metode evaluasi yang digunakan untuk mengukur tingkat keakuratan suatu model peramalan dengan menghitung rata-rata dari selisih absolut antara nilai aktual dan nilai prediksi yang dihasilkan oleh model (Muharrom, 2023). Nilai MAE yang lebih kecil menunjukkan bahwa model memiliki performa yang lebih baik dalam memprediksi data aktual, karena selisih antara nilai yang diprediksi dan nilai sebenarnya semakin kecil. Sebaliknya, semakin besar nilai MAE, semakin tinggi tingkat kesalahan dalam peramalan, yang menunjukkan bahwa model kurang mampu memberikan prediksi yang akurat (Nurani et al., 2023).

2.14 Mean Square Error (MSE)

Mean Squared Error (MSE) adalah metrik evaluasi yang umum digunakan dalam statistik dan machine learning untuk mengukur seberapa akurat sebuah model regresi dalam memprediksi nilai numerik (Nuha, 2023). MSE dihitung dengan cara mengambil rata-rata dari kuadrat selisih antara nilai prediksi dan nilai aktual. Dengan mengkuadratkan selisih, MSE memastikan bahwa semua kesalahan dihitung sebagai nilai positif, sehingga

memberikan gambaran yang lebih akurat tentang seberapa jauh prediksi menyimpang dari nilai sebenarnya

2.15 R²

Koefisien determinasi (R²) adalah metrik yang digunakan untuk mengukur seberapa baik model regresi menjelaskan variabilitas dalam data dependen. Nilai R² berkisar antara 0 hingga 1; semakin mendekati 1, semakin baik model dalam menjelaskan variabilitas data (Wijaya, n.d.).

Nilai R² berkisar antara 0 hingga 1:

R² = 0, berarti model tidak mampu menjelaskan variabilitas data sama sekali.

R² = 1, berarti model sepenuhnya mampu menjelaskan variabilitas data (model sempurna, yang sangat jarang terjadi).

0 < R² < 1, dengan kata lain model mampu menjelaskan sebagian variabilitas data, tetapi ada faktor lain yang belum dijelaskan oleh model (Ihzaniah et al., 2023).

2.16 Studi Literatur

Tabel 2. 1 Studi literatur

No	Judul	Permasalahan	Metode	Hasil	Keterbatasan
1	Penerapan Aplikasi RapidMiner Pada Prediksi Data Penjualan Album K-Pop Menggunakan Metode	Penjualan album yang tidak menentu menyebabkan kelebihan persediaan yang tidak terjual sehingga Perusahaan	Regresi Linier	Nilai RMSE sebesar 276051.973 dan nilai Absolute Error sebesar 97012.624. Dengan kata lain perhitungan RSME lebih besar dan lebih akurat.	Penelitian hanya menggunakan satu algoritma

	Regresi Linear	mengalami kerugian			
2	Implementasi Metode K-Nearest Neighbor dan Regresi Linear Dalam Prediksi Harga Emas	Fluktuasi harga emas yang terjadi setiap saat menjadi salah satu risiko yang umum dari investasi emas	Regresi Linier dan K-Nearest Neighbor	Metode Regresi Linier dan KNN menghasilkan nilai RMSE 0.05807 yang menunjukkan hasil akurasi yang baik.	Tidak ada pertimbangan factor eksternal dalam penelitian
3	Implementasi Algoritma Regresi Linier dan K-Nearest Neighbor Untuk Prediksi Harga Rumah	Pentingnya pemahaman mendalam mengenai factor-faktor yang mempengaruhi nilai property guna memutuskan pembelian property.	Regresi Linier dan K-Nearest Neighbor	Nilai relative error regresi linier sebesar 71.95% +/- 82.63%. Sedangkan hasil relatif error K-Nearest Neighbor sebesar 30.95% +/- 29.87%. Dengan kata lain nilai perbandingan Tingkat akurasi menggunakan algoritma KNN lebih baik daripada Regresi Linier.	Tidak adanya Batasan masalah dari penelitian ini.
4	Analisis Algoritma Regresi Linear Sederhana dalam Memprediksi Tingkat Penjualan	Tren penjualan album yang tidak menentu terlebih pada grup NCT 127 yang telah menemukan	Regresi Linear	Hasil penelitian menggunakan algoritma regresi linear sederhana 10.86%, serta hasil performance vector dengan rapidminer sebesar 0.089%	Penelitian hanya menggunakan satu algoritma

	Album KPOP	kresuksesan internasional			
5	Algoritma K-Nearest Neighbor Classification sebagai Sistem Prediksi Predikat Prestasi Mahasiswa	Bertambahnya mahasiswa menimbulkan pola yang bervariasi pada atribut data mahasiswa serta berpotensi melakukan data mining yang akan memberikan pengetahuan baru	Regresi Linear	Perhitungan algoritma K-NN dengan perhitungan confusion matrix yang membandingkan predikat pada semester sebelumnya dengan predikat hasil prediksi menghasilkan akurasi sebesar 82%	Perlunya pemantauan lebih lanjut mengenai kemajuan predikat prestasi mahasiswa lebih lanjut
6	Implementasi Machine Learning untuk Prediksi Harga Mobil Bekas dengan Algoritma Regresi Linear Berbasis Web	Banyaknya factor penentu dari penentuan harga mobil bekas, seperti model, tahun produksi, yang menjadi kendala dalam pengambilan Keputusan saat akan membeli atau menjual mobil	Regresi Linier	Hasil prediksi mencapai akurasi sebesar 76.22% serta menghasilkan nilai RMSE 3813.65 setelah melakukan pengujian sebanyak 10 kali menggunakan metode black box	Keterbatasan variable mempengaruhi hasil akurasi
7	Prediksi Penjualan	Dinamika dalam	K-Nearest	Dengan pemanfaatan	Penelitian hanya

	Obat dan Alat Kesehatan Terlaris Menggunakan Algoritma K-Nearest Neighbor	manajemen ketersediaan obat dan alat Kesehatan menjadi hal krusial dalam memberikan layanan Kesehatan yang optimal untuk pasien	Neighbor	aplikasi Jupyter Notebook yang kemudian diuji nilai akurasi menggunakan confusion matrix, diketahui nilai accuracy 1.0, recall 1.0, dan precision 1.0.	menggunakan satu algoritma
8	Sistem Prediksi Jumlah Obat Keluar Menggunakan Metode Regresi Linier untuk Menentukan Jumlah Pemesanan Obat	Pelayanan Kesehatan yang tidak maksimal dan menyulitkan pasien untuk mendapat obat sehingga harus membeli obat di apotek lain menjadi akibat dari kebutuhan obat yang tidak tersedia	Regresi Linier	Nilai perhitungan MAPE pada prediksi penelitian ini adalah 12.42%. Berdasarkan perhitungan hasil kuesioner menggunakan UAT didapat hasil 74.64%.	Penelitian hanya menggunakan satu algoritma
9	Penerapan K-Nearest Neighbor Classification dalam Penentuan Status Gizi Ibu Hamil	Guna menghindari berbagai risiko akibat kekurangan atau kelebihan gizi, ibu	Regresi K-Nearest Neighbor	Dengan perhitungan Sistem Pendukung Keputusan (SPK) yang membandingkan antara jumlah	Variabel inputan dipilih secara manual yang menghabis

	Berdasarkan Tingkat Konsumsi Makanan Secara Harian	hamil harus memperhatikan bahan makanan yang dikonsumsi untuk mendukung pertumbuhan janin dan Kesehatan ibu		pengujian yang benar dan jumlah data sampel, diperoleh persentase 86.7%	kan banyak waktu
10	Penerapan Data Mining untuk Prediksi Penjualan Produk Elektronik Terlaris Menggunakan Metode K-Nearest Neighbor	Dibutuhkan prediksi untuk penjualan elektronik yang paling laris akibat banyaknya permintaan konsumen akan produk elektronik	K-Nearest Neighbor	Melalui tahapan Knowledge Discovery in Database (KDD), diketahui nilai akurasi terhadap klasifikasi penjualan produk terlaris sebesar 92.51%	Penelitian hanya menggunakan satu algoritma

