



6.2%

SIMILARITY OVERALL

SCANNED ON: 7 JUL 2024, 6:36 PM

Similarity report

Your text is highlighted according to the matched content in the results above.

IDENTICAL 0.23% **CHANGED TEXT** 5.97% **QUOTES** 1.15%

Report #21945037

1 BAB I PENDAHULUAN 1.1 Latar Belakang Seiring dengan kemajuan teknologi dan ekspansi global, pendidikan di Indonesia saat ini sedang mengalami transformasi yang signifikan. Pendidikan merupakan salah satu bidang kehidupan yang terkena dampak dari era informasi dan digitalisasi. Sebagai komponen penting dalam kerangka pendidikan di Indonesia, pendidikan tinggi sangat penting untuk mengembangkan tenaga kerja terampil. Meski mengalami kemajuan yang pesat, pendidikan tinggi di Indonesia masih dihadapkan pada sejumlah tantangan, seperti ketimpangan kualitas antar wilayah, keterbatasan akses ke sumber daya pendidikan yang bermutu, serta kebutuhan penyesuaian kurikulum dengan tuntutan industri 4.0. Pendidikan tinggi tidak hanya bertujuan untuk meningkatkan kualitas individu, tetapi juga diharapkan menjadi pendorong dalam pembangunan ekonomi dan sosial bangsa. (Moh Abdul, Elin, dkk. 2022). Berdasarkan Undang-Undang Republik Indonesia No 37 12 Tahun 2012 Tentang Pendidikan Tinggi bahwa 2 “Pendidikan tinggi sebagai bagian dari sistem pendidikan nasional memiliki peran strategis dalam mencerdaskan kehidupan bangsa dan memajukan ilmu pengetahuan dan teknologi dengan memperhatikan dan menerapkan nilai humaniora serta kebudayaan dan pemberdayaan bangsa Indonesia yang berkelanjutan 37 . Persaingan antar perguruan tinggi menjadi semakin ketat, baik dalam hal menarik mahasiswa yang berkualitas maupun dalam menghasilkan lulusan yang siap memasuki pasar kerja yang dinamis dan kompetitif. Oleh karena itu, manajemen

universitas yang ketat dalam hal pengajaran, penelitian, dan pengabdian masyarakat menjadi sangat penting. Strategi ini harus mampu mengatasi tantangan era modern serta memenuhi harapan dari para pemangku kepentingan, termasuk mahasiswa, industri, dan pemerintah. Pendaftaran mahasiswa baru merupakan salah satu aspek kunci dalam pengelolaan perguruan tinggi. Proses ini tidak hanya tentang merekrut calon mahasiswa, tetapi juga tentang bagaimana menarik individu-individu yang sesuai dengan nilai dan visi universitas. Ketika memilih universitas, calon mahasiswa terkadang mempertimbangkan faktor-faktor seperti biaya, prospek kerja, reputasi institusi, dan kualitas program studi. Dalam iklim yang semakin kompetitif, sangat penting untuk memahami dinamika pendaftaran mahasiswa baru. Universitas perlu mengidentifikasi elemen-elemen yang mempengaruhi keputusan calon mahasiswa dan menyesuaikan pendekatan mereka. Salah satu aspek krusial dalam pendaftaran mahasiswa baru adalah memahami minat calon mahasiswa terhadap program studi yang tersedia. Pemahaman tentang pola minat ini penting bagi universitas untuk dapat mengatur strategi pengenalan program studi yang efektif. Selain itu, analisis data minat ini juga penting untuk melihat gambaran yang lebih jelas tentang kebutuhan mahasiswa, yang pada gilirannya akan membantu universitas dalam merencanakan dan mengimplementasikan strategi penerimaan dan pengembangan akademik yang lebih efektif. Penelitian ini



dilatarbelakangi oleh serangkaian temuan yang muncul dalam pengelolaan penerimaan mahasiswa baru di Universitas XYZ. Belum pernah dilakukan analisis data secara sistematis terkait dengan penerimaan mahasiswa baru, sehingga pemahaman yang mendalam atau 2 berbasis data mengenai pola penerimaan mahasiswa sebelumnya belum terbentuk. Selain itu, terdapat kekurangan informasi tentang minat calon mahasiswa terhadap program studi yang mereka pilih. Tantangan tambahan muncul dalam mengelola pemasaran yang efektif dan berbasis data untuk menjangkau calon mahasiswa yang sepaham dengan visi dan misi Universitas XYZ. Keterbatasan dalam mengembangkan program studi yang responsif terhadap kebutuhan industri juga menjadi perhatian utama. Dalam hal ini, analisis data menjadi kunci dalam memahami dan merespons kebutuhan serta minat calon mahasiswa. Data yang terkumpul dari proses pendaftaran mahasiswa baru dapat memberikan wawasan berharga tentang preferensi yang berlaku. Dengan mengaplikasikan metode analisis K-Nearest Neighbor (K-NN) yang dipilih dalam studi kasus ini, universitas dapat mengenali pola-pola yang mungkin tidak tampak secara langsung. Metode ini dipilih karena kemampuannya dalam melakukan klasifikasi dan prediksi yang lebih akurat terkait dengan minat dan kecenderungan calon mahasiswa. K-NN lebih fleksibel dalam mengelola data yang rumit dan tidak terstruktur daripada teknik-teknik yang sebanding seperti Pohon Keputusan atau Naive Bayes. Selain itu, K-NN juga tidak

membuat asumsi tentang distribusi data, sehingga cocok digunakan untuk analisis pola minat calon mahasiswa yang mungkin bervariasi. Penelitian ini diperkirakan akan memberikan beberapa manfaat penting. Pertama, hasil analisis ini bisa membantu universitas dalam merancang strategi pemasaran dan promosi yang lebih efisien untuk menarik calon mahasiswa. Kedua, penelitian ini dapat menawarkan perspektif yang berharga untuk menciptakan kurikulum yang selaras dengan minat calon mahasiswa, sehingga meningkatkan kualitas dan penerapan pendidikan yang diberikan. Ketiga, dengan mempelajari minat dan kebutuhan calon mahasiswa, perguruan tinggi dapat memodifikasi kurikulum dan metode pengajaran mereka untuk membuat ruang kelas lebih menarik dan produktif. Implikasi jangka panjang dari penelitian ini sangatlah signifikan. Melalui pemahaman yang lebih baik tentang dinamika pendaftaran mahasiswa baru dan minat calon mahasiswa, perguruan tinggi dapat mengambil langkah-langkah strategis dalam menghadapi persaingan dan tantangan di masa depan. Penelitian ini juga dapat memberikan kontribusi dalam pengambilan keputusan strategis di tingkat kebijakan pendidikan tinggi, khususnya dalam hal pengembangan program studi dan alokasi sumber daya. Penelitian ini mencerminkan kebutuhan penting untuk memahami kompleksitas dan dinamika dalam penerimaan mahasiswa baru di perguruan tinggi Indonesia. Melalui penggunaan metode K-Nearest Neighbor dalam analisis data, penelitian ini tidak hanya memberikan wawasan

penting bagi Universitas XYZ, tetapi juga memberikan kontribusi lebih luas terhadap pemahaman tentang bagaimana perguruan tinggi dapat menavigasi dan merespons tantangan serta peluang dalam lingkungan pendidikan yang terus berubah. Dengan fokus pada analisis data dan pemahaman minat calon mahasiswa, penelitian ini memberikan panduan berharga dalam pengembangan strategi penerimaan dan akademis yang efektif untuk masa depan pendidikan tinggi di Indonesia. Peneliti mengidentifikasi penelitian ini dengan judul tersebut setelah mempertimbangkan penjelasan yang diberikan dalam deskripsi yaitu, “Penerapan Data Mining Untuk Memprediksi Peminatan Program Studi Dengan Metode K-Nearest Neighbor (Studi Kasus : Universitas XYZ)

3 1.2 Identifikasi dan Rumusan Masalah Informasi latar belakang di atas mengarah pada identifikasi beberapa masalah, termasuk: 1. Belum pernah dilakukannya analisis data secara sistematis terkait penerimaan mahasiswa baru. Sehingga, belum ada pemahaman yang mendalam atau berbasis data mengenai pola penerimaan mahasiswa sebelumnya 2. Belum adanya informasi yang cukup tentang minat calon mahasiswa terhadap program studi yang mereka pilih. Oleh karenanya, kemudian peneliti mengembangkan isu yang akan dibahas dalam investigasi ini, menjadi: ❑ Bagaimana akurasi i prediksi minat calon mahasiswa terhadap program studi pilihan di waktu depan di Universitas XYZ? 1.3 Ruang Lingkup dan Batasan Masalah Poin-poin berikut ini dapat digunakan untuk mengkarakterisasi cakupan. 1. Penelitian ini menggunakan Data Penerimaan Mahasiswa Baru Universitas XYZ 2. Periode waktu analisis sendiri menggunakan data dari Penerimaan Mahasiswa Baru yang dimulai dari tahun 2020 hingga 2023. 3. Penelitian akan menggunakan metode K-NN untuk menganalisis data. Ini akan mencakup pengembangan model, pengujian, dan validasi model untuk memastikan keakuratan dalam prediksi. 1.4 Tujuan Penelitian Tujuan dari penelitian ini adalah untuk memberikan pemahaman yang komprehensif mengenai tingkat ketertarikan calon mahasiswa terhadap program-program akademik yang ditawarkan oleh Universitas XYZ, memprediksi akurasi minat ini di masa mendatang, dan memanfaatkan hasilnya untuk merancang strategi pengenalan

universitas di kalangan siswa sekolah menengah atas. Tujuan ini mencakup pembuatan materi informasi yang ditargetkan untuk sekolah menengah atas guna meningkatkan kesadaran dan minat terhadap program-program universitas, serta menyusun rencana untuk memperkenalkan universitas yang tidak hanya bertujuan untuk meningkatkan jumlah mahasiswa baru yang mendaftar di Universitas XYZ, tetapi juga untuk meningkatkan profil institusi dan membuatnya lebih menarik. 10 21 1.5 Manfaat Penelitian Penelitian ini diharapkan dapat memberikan beberapa manfaat nyata bagi para pemangku kepentingan.

Manfaat-manfaat tersebut seperti yang dijelaskan di bawah ini: A.

Untuk Peneliti - Penelitian ini menawarkan kesempatan kepada para peneliti untuk meningkatkan dan memperdalam keterampilan analitis mereka, terutama dalam menerapkan metode K-Nearest Neighbor pada data pendidikan.

- Berkontribusi pada literatur akademik, khususnya dalam studi tentang analisis data pendidikan, dan dapat menjadi fondasi bagi penelitian

lanjutan. 4 - Mendapatkan pemahaman yang lebih dalam tentang proses penerimaan mahasiswa baru, prediksi minat studi, dan dinamika pendidikan

tinggi. B. Untuk Pembaca Memberikan gambaran terkini dalam pendidikan

tinggi, terutama terkait dengan minat dan preferensi calon mahasiswa, yang dapat menjadi informasi berharga bagi pembaca serta dapat menjadi

referensi penting bagi peneliti lain yang ingin melakukan studi serupa

atau terkait di bidang pendidikan tinggi. C. Untuk Institusi - Membantu institusi untuk mengetahui akurasi prediksi pendaftaran mahasiswa baru,

memungkinkan identifikasi pola dan preferensi calon mahasiswa. -

Mendapatkan keuntungan dari kemampuan untuk membuat keputusan yang lebih berbasis data dan strategis. - Temuan penelitian akan menyediakan

informasi yang dapat membantu dalam pengembangan atau penyesuaian program studi di universitas, menjadikannya lebih relevan dan menarik bagi calon

mahasiswa. 1.6 Sistematika Penulisan Dalam sistematika, dijelaskan apa yang terdapat di setiap bab dengan susunan yang sistematis sehingga pemahaman

tentang penelitian dapat disampaikan dengan jelas. BAB I PENDAHULUAN Pada

BAB I, latar belakang penelitian memulai pembahasan dengan menyoroti

pentingnya analisis data dalam proses penerimaan mahasiswa baru di Universitas XYZ. Dalam bab ini, juga dipaparkan rumusan masalah yang menjelaskan tantangan spesifik yang dihadapi universitas, serta tujuan penelitian yang diarahkan untuk mengatasi tantangan tersebut. Manfaat penelitian dijelaskan dengan fokus pada nilai tambah yang diberikan kepada peneliti, institusi, dan pembaca. Definisi ruang lingkup penelitian digunakan untuk menetapkan batasan penelitian, dan bab ini diakhiri dengan penjelasan sistematika penulisan yang menguraikan struktur keseluruhan penelitian.

16 BAB II TINJAUAN PUSTAKA BAB II berfokus pada pengkajian literatur yang relevan, memberikan dasar teori dan konseptual untuk penelitian. Bab ini mengulas konsep-konsep terkait penerimaan mahasiswa baru, metodologi analisis data seperti K-Nearest Neighbor, dan studi terkait yang telah dilakukan sebelumnya. Tinjauan pustaka ini bertujuan untuk memberikan konteks dan pemahaman teoretis yang mendalam tentang topik penelitian, serta menetapkan fondasi untuk metodologi penelitian. **15** BAB III METODE PENELITIAN Metode Penelitian, yang merupakan bab ketiga, memberikan gambaran tentang teknik metodologi yang digunakan dalam penelitian. Bab ini membahas tentang desain penelitian, metode pengumpulan data (seperti bagaimana memilih sampel dan mengumpulkan data), dan metode analisis data yang digunakan untuk mencari tahu apa arti dari data tersebut. Relevansi bab ini terletak pada fakta bahwa 5 bab ini menjamin bahwa penelitian akan dilakukan secara metodis dan dapat dipertanggung jawabkan. BAB IV HASIL DAN PEMBAHASAN BAB IV memuat temuan dari analisis data beserta pembahasan yang mendetail mengenai temuan tersebut. Bab ini mengartikan hasil dalam konteks rumusan masalah dan tujuan penelitian, menganalisis data dengan merujuk pada teori dan penelitian sebelumnya, serta membahas implikasi temuan terhadap praktik dan teori yang telah ada. BAB V PENUTUP BAB V Penutup menyimpulkan keseluruhan penelitian, merangkum temuan utama, dan menggarisbawahi bagaimana penelitian ini menjawab rumusan masalah. Bab ini juga menyediakan rekomendasi berdasarkan hasil penelitian, baik untuk praktik di Universitas XYZ maupun untuk penelitian masa depan. Selain

itu, bab ini mengakui keterbatasan penelitian, memberikan perspektif tentang ruang lingkup di mana temuan dapat diinterpretasikan dan diterapkan. 10 36 6

BAB II TINJAUAN PUSTAKA 2.1 Teori Dasar 2.1 1Perguruan Tinggi Berdasarkan Peraturan Pemerintah Republik Indonesia Nomor 30 Tahun 1990 Tentang Pendidikan Tinggi bahwa “Pendidikan tinggi adalah satuan pendidikan yang menyelenggarakan pendidikan tinggi yang merupakan jenjang yang lebih tinggi daripada pendidikan menengah di jalur pendidikan sekolah . Institusi perguruan tinggi menyediakan berbagai program studi dan penelitian di berbagai disiplin ilmu. Fakultas-fakultas di perguruan tinggi biasanya menawarkan program sarjana, magister, dan doktoral dalam bidang-bidang seperti ilmu sosial, ilmu alam, teknik, kedokteran, dan humaniora. Selain itu, perguruan tinggi berperan penting dalam pengembangan penelitian ilmiah dan kontribusi terhadap kemajuan ilmu pengetahuan dan teknologi. Di banyak kasus, perguruan tinggi juga menjadi tempat untuk kegiatan ekstrakurikuler, aktivitas sosial, dan pengembangan kepribadian mahasiswa. 13 Selain itu, di dalam Peraturan Pemerintah Republik Indonesia Nomor 30 Tahun 1990 Tentang Pendidikan Tinggi juga memuat tujuan pendidikan tinggi, yaitu: - 1 “Menyiapkan peserta didik menjadi anggota masyarakat yang memiliki kemampuan akademik dan /atau profesional yang dapat menerapkan, mengembangkan dan/atau menciptakan ilmu pengetahuan, teknologi dan/atau kesenian 13 . - 1 “Mengembangkan dan menyebarkan ilmu pengetahuan, teknologi dan/atau kesenian serta mengupayakan penggunaannya untuk meningkatkan taraf kehidupan masyarakat dan memperkaya kebudayaan nasional 13 . 2.1.2Penerimaan Mahasiswa Baru Mahasiswa diterima di perguruan tinggi negeri dan swasta melalui proses Penerimaan Mahasiswa Baru (PMB) yang dilakukan. Proses ini melibatkan seleksi oleh perguruan tinggi atau universitas untuk menerima calon mahasiswa ke dalam program studi yang tersedia. Setiap tahun, perguruan tinggi melaksanakan penerimaan mahasiswa baru ini yang umumnya melibatkan beberapa tahapan seperti pengumpulan berkas pendaftaran, ujian tertulis, wawancara, dan evaluasi prestasi akademik. Biasanya, calon mahasiswa baru mendaftar secara online atau offline sesuai dengan prosedur yang berlaku.

Mereka juga diminta untuk menyerahkan dokumen seperti transkrip nilai, surat rekomendasi, dan hasil ujian tertentu. (P. Andriani, 2019).

2.1.3 Data Data adalah pengukuran atau pengamatan yang dikumpulkan sebagai sumber informasi. Data mengacu pada seperangkat nilai yang pada variabel (apa yang diukur) dan unit pengamatan (anggota sampel atau populasi). Data adalah bagian penting dalam statistik, yang perlu dipahami untuk diterapkan pengukuran statistik dengan benar dalam menyimpulkan asumsi tertentu dengan benar. Data mempunyai peran penting dalam bidang statistik. Pengetahuan dan pemahaman tentang data akan membantu dalam menerapkan 7 pengukuran statistik secara akurat dari data mentah untuk menyimpulkan hasil yang bermakna (Handayani, 2022). Jenis data yang berbeda merepresentasikan data yang berbeda juga. Ada dua jenis data berdasarkan sifat diantaranya. 1) Data Kualitatif Data yang disajikan berbentuk verbal atau naratif merupakan data kualitatif yang dikumpulkan melalui diskusi kelompok, wawancara, kuesioner dengan pertanyaan terbuka, serta situasi lain yang kurang terstruktur. Data kualitatif umumnya memiliki sifat objektif, sehingga dapat menghasilkan berbagai interpretasi dari setiap pembacanya. 2) Data Kuantitatif Data kuantitatif merupakan data untuk direpresentasikan berbentuk angka, di mana angka-angka tersebut bisa bervariasi besar atau kecil. Angka-angka ini mungkin sesuai dengan kategori atau label tertentu. Data kuantitatif bisa didapatkan melalui survei yang menghasilkan jawaban kaku berupa angka. Contohnya, usia Bambang adalah 21 tahun, dan tinggi rata-rata siswa di kelas A adalah 170 cm. 2.1.4 Basis Data Basis data adalah sebuah lokasi di dalam sebuah organisasi atau bisnis di mana data dihubungkan untuk memfasilitasi pencarian dan penggunaan kembali data yang lebih cepat dan mudah. Sebaliknya, sistem basis data menggunakan komputer untuk mengelola dan mengatur catatan untuk menyimpan dan melestarikan semua data yang tersedia dalam organisasi atau bisnis, memberikan informasi terbaik bagi pengguna untuk mengambil keputusan (Adyanata Lubis, 2016). Berikut ini adalah beberapa bagian yang membentuk sistem basis data: 1) Hardware

Perangkat komputer standar yang mendukung pemrosesan data adalah perangkat keras yang dimaksud. Di dalam sistem basis data, perangkat keras yang sering digunakan meliputi server database, hard disk, dan komponen lainnya. 20 2) Operating System Sistem operasi, manajemen sumber daya, dan eksekusi program semuanya dilakukan oleh program pada komputer. 3) Database Management System (DBMS) Pengguna tidak langsung melakukan operasi fisik pada basis data, tetapi mengatur penyimpanan, modifikasi, dan pengambilan data menggunakan program yang dikenal sebagai sistem manajemen basis data (DBMS). 4) Pengguna (User) Pengguna menggunakan program yang ditulis dalam bahasa pemrograman untuk berkomunikasi dengan basis data dan mengelola data. 5) Aplikasi atau perangkat lain Aplikasi ini bergantung pada kebutuhan, di mana pengguna basis data dapat membuat program khusus yang mudah digunakan untuk memasukkan, mengubah, dan mengambil data. Program-program tersebut dapat dijalankan langsung dengan DBMS atau ditulis menggunakan bahasa pemrograman. 8 2.1.5 Algoritma Algoritma merupakan metode efektif yang direpresentasikan sebagai serangkaian instruksi terbatas yang telah dijelaskan secara jelas untuk menyelesaikan suatu fungsi. Algoritma pemecahan masalah, sederhananya, adalah serangkaian tahapan logis. Ketika dihadapkan pada suatu masalah, seseorang harus melakukan pendekatan terhadap solusinya secara logis. Prosedur langkah demi langkah yang digunakan untuk perhitungan dikenal sebagai algoritme dalam disiplin ilmu komputer dan matematika (Sigit Susanto, dkk., 2022). Urutan algoritma dimulai dengan kondisi atau input awal, diikuti oleh instruksi- instruksi yang menjelaskan komputasi. Saat dieksekusi dan diproses melalui urutan kondisi yang telah didefinisikan dengan baik, ini menghasilkan output dan berakhir pada kondisi akhir yang ditentukan. Dalam hal pemrosesan data komputer, algoritme sangat penting. Insinyur perangkat lunak menggunakan algoritme dalam sistem komputer untuk membuat logika yang membuat perangkat lunak lebih mudah digunakan. Ada tiga jenis algoritma yang mendasar, yaitu. 1) Sequence Algorithm Algoritme ini mengikuti serangkaian langkah yang harus diselesaikan secara berurutan. Mendidihkan air adalah

ilustrasi langsung dari algoritme ini, di mana setiap langkah harus diselesaikan secara berurutan, mulai dari atas dan terus ke bawah. 2) Looping Algorithm Algoritme yang mengulang serangkaian tahapan yang telah ditentukan sebelumnya. Dalam beberapa situasi, ada kebutuhan untuk melakukan langkah- langkah tertentu berulang kali. 3) Conditional Algorithm Algoritme ini melanjutkan ke fase berikutnya jika kondisi yang telah ditentukan terpenuhi. 2.1 **2** **19** Knowledge Discovery in Database (KDD) Proses menemukan pengetahuan dalam basis data disebut dengan Knowledge Discovery in Database (KDD). **9** Proses penggalian atau pengenalan pola, pengetahuan, dan informasi yang berpotensi berguna dari kumpulan data yang sangat besar tercakup dalam deskripsi KDD yang komprehensif. Pengetahuan dan data yang dihasilkan oleh KDD harus dapat diandalkan, orisinal, dapat dipahami, dan praktis. Proses menemukan data yang valid, unik, mungkin berguna, dan akhirnya memiliki pola yang masuk akal disebut sebagai penemuan pengetahuan dalam database. **38** Ini bukanlah tugas yang mudah untuk dicapai. Mengintegrasikan dan menavigasi berbagai sumber data merupakan masalah umum dalam penemuan pengetahuan, terutama ketika data menjadi lebih multidimensi dan membutuhkan metode baru untuk pengenalan pola. Penemuan pengetahuan dalam data, atau KDD, didefinisikan sebagai proses mengeksplorasi dan menganalisis data dalam jumlah besar untuk mendapatkan pengetahuan dan informasi yang bermanfaat (Fitri Marisa et al., 2021). 9 Gambar 2. 1 Tahapan KDD Sumber : Han & Kamber (2006) Gambar 2.1 mengilustrasikan fase-fase utama dari proses KDD, yang meliputi tahapan-tahapan berikut: 1) Data Cleaning Menghilangkan duplikasi, memastikan konsistensi data, dan memperbaiki kesalahan pengetikan adalah bagian dari proses pembersihan data. **2** Secara umum, data yang diperoleh dari hasil uji coba atau database perusahaan sering kali tidak lengkap, mengandung kekurangan seperti data yang hilang atau tidak akurat atau kesalahan ketik. Selain itu, terkadang terdapat properti data yang tidak sesuai dengan teori data mining yang sedang diteliti. Dengan semakin sedikitnya data yang diproses dan semakin sedikitnya data yang kompleks yang ditangani, prosedur pembersihan data ini juga dapat



berdampak pada output informasi dari teknik data mining. 2) Data Integration Untuk membuat satu database baru yang dibutuhkan KDD, data harus diintegrasikan dari beberapa database atau digabungkan dengan informasi terkait lainnya. Integrator data harus menggabungkan beberapa set data dan menghilangkan noise dari sumber data untuk menyelesaikan tahap pembersihan dan integrasi KDD. 3) Data Selection Metode memilih informasi yang relevan untuk diperiksa dari data operasional. Basis data lain berisi data yang dipilih. 4) Data Transformation Yaitu, mengubah format data ke dalam bentuk tertentu yang dapat digunakan oleh metode data mining. Beberapa teknik umum tertentu, seperti analisis asosiasi dan pengelompokan, membutuhkan data masukan dalam format kategorikal. 5) Data Mining Menggunakan strategi, taktik, atau algoritme tertentu untuk mencari tren atau data yang menarik. 6) Pattern Evaluation Mengenali tren yang menarik dalam temuan data mining. Pada titik ini, keberhasilan hipotesis saat ini ditentukan dengan menganalisis output dari teknik penggalian data seperti membedakan pola atau model prediksi. 7) Knowledge Presentation Menampilkan pola informasi yang dihasilkan oleh metode data mining. Grafik ini memfasilitasi komunikasi yang jelas dan sederhana dari hasil data mining. 10 2.1.7 Data Warehouse Menurut Bill Inmon, yang diakui sebagai bapak Data Warehouse, Sebuah "gudang data" adalah sekelompok kumpulan data yang berorientasi pada subjek, gabungan, tidak berubah-ubah, dan varian waktu yang membantu manajer mengambil keputusan. 3 Definisi lain yang terkait dengan Business Intelligence (BI) menyatakan bahwa data warehouse adalah basis data yang didesain untuk digunakan dalam aktivitas BI, membantu pengguna dalam memahami dan meningkatkan kinerja organisasi mereka (Lane dan Potineni, 2014). Data warehouse seringkali dikaitkan dengan BI karena menjadi sumber data untuk analisis bisnis. Fokus data warehouse adalah menyediakan data untuk analisis bisnis, bukan untuk proses transaksi atau operasional bisnis. Data warehouse juga sering disebut sebagai database Online Analytical Processing (OLAP) karena perannya dalam memenuhi kebutuhan analitik (Wardani, dkk., 2019). Data

Warehouse merupakan sumber daya yang memerlukan pengolahan untuk menghasilkan informasi, baik berupa data kualitatif maupun kuantitatif, yang menggambarkan fakta dan memberikan manfaat bagi peneliti atau memberikan gambaran tentang kondisi atau situasi tertentu. Informasi merupakan data yang telah diproses untuk menghasilkan analisis yang diperlukan oleh pengguna (Gunadi & Widiyanto, 2020). Definisi data warehouse berdasarkan fungsinya dalam menyediakan informasi menurut Ponniah (2010) adalah sebagai berikut: 1. Memberikan perspektif yang komprehensif dan terintegrasi terhadap informasi bisnis atau organisasi. 2. Menyediakan akses mudah ke informasi terkini dan historis untuk mendukung pengambilan keputusan strategis. 3. Membuat informasi dari data operasional tersedia tanpa mengganggu sistem operasional utama. 4. Mengintegrasikan informasi dari organisasi atau perusahaan secara konsisten. **31** 5. Menyajikan sumber informasi secara strategis dengan cara yang fleksibel dan interaktif. **8** Keempat karakteristik data warehouse (terintegrasi, berbasis waktu, berorientasi pada subjek, dan tidak mudah berubah) tetap menjadi panduan dalam pengembangan data warehouse. Berorientasi pada subjek menyiratkan bahwa data warehouse menyajikan pandangan yang terfokus dan jelas mengenai masalah subjek tertentu dengan mengabaikan data yang tidak relevan dalam proses pengambilan keputusan (Han, Kamber, dan Pei, 2012). Salah satu pendekatan dalam membangun data warehouse berorientasi pada subjek adalah penggunaan data warehouse untuk memberikan informasi tentang kinerja dosen. Dalam hal ini, dosen adalah subjeknya. Ini berbeda dengan data operasional sehari-hari atau aktivitas dosen yang tercatat di berbagai aplikasi seperti Sistem Informasi Perkuliahan, Sistem Informasi Bimbingan Skripsi, dan Sistem Informasi Kehadiran Pegawai berbasis OLTP (Online Transactional Processing). **22** Database OLTP dapat berfungsi sebagai sumber data untuk membangun data warehouse berbasis OLAP (Online Analytical Processing). **11** 2.1.8 Data Mining Teknik mengekstraksi pola dan informasi yang relevan dari data dalam jumlah besar dikenal sebagai data mining. Ekstraksi data, analisis, analisis, dan pengumpulan data adalah bagian dari proses ini. Analisis

matematis digunakan untuk mengidentifikasi pola dan tren dalam data yang sering kali sulit atau rumit untuk diidentifikasi melalui metode eksplorasi data tradisional, sebagai akibat dari hubungan yang rumit antara data atau volume data yang besar. Menemukan pola yang sebelumnya tidak teridentifikasi adalah tujuan utama dari data mining, yang kemudian dapat diterapkan pada penyelesaian berbagai masalah (M. Arhami & M. Nasir, 2020). Data mining telah menjadi teknologi yang kuat dengan potensi besar untuk membantu perusahaan mengidentifikasi informasi penting dari data yang mereka kumpulkan mengenai perilaku pelanggan dan prospek. Informasi yang dihasilkan dari data mining dapat digunakan sebagai alat bantu pengambilan keputusan untuk mengembangkan bisnis. Data mining mampu memberikan solusi atas berbagai pertanyaan bisnis yang kompleks. Selain itu, data mining juga dapat digunakan untuk membuat prediksi di masa depan, sehingga pengusaha dapat mengambil keputusan yang efektif, proaktif, dan dinamis. Proses data mining secara umum meliputi tiga tahap: 1. Eksplorasi Prosesnya dimulai dengan menyiapkan banyak data, yang kemudian dibersihkan atau dikurangi sesuai kebutuhan, dan materi duplikat dihapus untuk menyisakan hanya data yang berguna. 2. Pemodelan atau identifikasi pola Tahap selanjutnya adalah membangun model statistik yang bertujuan untuk menilai dan menghasilkan prediksi yang paling akurat dan terbaik. Karena perlunya pengujian dan perbandingan temuan yang sering dilakukan, prosedur ini dapat memakan waktu jika menggunakan kumpulan data yang sama untuk model. 3. Penerapan Dengan menggunakan data pelatihan dan pengujian, model dievaluasi pada langkah terakhir untuk menghasilkan estimasi atau prakiraan hasil yang diinginkan. Alur proses data mining dijelaskan oleh arsitektur data mining, yang mencakup segala sesuatu mulai dari administrasi sumber data hingga interaksi antara pengguna dan sistem data mining. Gambar 2.2 menampilkan arsitektur untuk data mining. Gambar 2. 2 Arsitektur Data Mining Sumber : Indriani (2023) Menurut Efori Buulolo (2020), sumber data untuk data mining dapat ditemukan di berbagai lokasi, termasuk database, gudang, internet, dan sumber lainnya.

Setelah data yang tidak berguna dihilangkan, data dari beberapa sumber ini digabungkan dan dipilih. Selanjutnya, data yang dipilih diformat sedemikian rupa sehingga dapat digunakan untuk 12 penggalian data. Data yang telah dikonversi kemudian diproses untuk menciptakan pengetahuan atau informasi yang dapat ditampilkan kepada pengguna. 2.1 7 9 Metode Klasifikasi

Klasifikasi adalah proses mengidentifikasi model atau fungsi yang mencirikan dan memisahkan kelas-kelas data atau konsep untuk memprediksi kelas data yang kelasnya tidak diketahui. Data pelatihan, atau objek data dengan label kelas yang telah ditentukan, dianalisis untuk membangun model. Model akhir dapat ditampilkan dalam berbagai cara, termasuk pohon keputusan dan klasifikasi IF-THEN. Data dikelompokkan menggunakan nilai atribut klasifikasi dan data pelatihan agar algoritme klasifikasi dapat berfungsi. 6 Data baru dikategorikan ke dalam kelompok yang sudah ada sebelumnya dengan menggunakan metode pengelompokan ini. Pohon keputusan adalah salah satu ilustrasi klasifikasi. 1 5 23 Dalam sebuah pohon keputusan, setiap node menunjukkan tes atribut dari kumpulan data, dan setiap cabang menunjukkan hasil tes. 1 5

6 Berdasarkan kondisi IF THEN outcome, pohon keputusan yang dibuat dapat dilihat sebagai sekumpulan aturan. 1 Berlawanan dengan label kelas, tujuan pengguna sering kali adalah untuk memperkirakan nilai dari data yang tidak tersedia atau hilang. Prediksi dalam situasi ini terutama berkaitan dengan mengidentifikasi distribusi dari data yang disediakan. (Amril Mutoi & Adam Puspabhuana, 2017). 2.1.10 K-Nearest Neighbor Salah satu pendekatan pembelajaran mesin dan penggalian data yang digunakan untuk regresi dan klasifikasi disebut K-nearest neighbor, atau K-NN. K-NN bekerja dengan melihat dataset dan mencari tahu kategori atau nilai tujuan dari data baru dengan melihat sebagian besar data terdekatnya. K-NN menghitung jarak antara data baru dan data lama selama proses berlangsung. Algoritme ini kemudian menggunakan sebagian besar tetangga terdekat dari data baru untuk mengetahui kategori atau nilai tujuannya. Ketika melakukan klasifikasi atau regresi, jumlah tetangga ditentukan oleh nilai K; nilai K yang lebih besar mengindikasikan lebih banyak tetangga

untuk digunakan. (Ferdina Kusumah, et al., 2022). Metode K-NN merupakan bagian dari Instance-Based Learning (IBL), yaitu teknik machine learning yang mempelajari saat sebuah masukan diberikan untuk klasifikasi. Berbeda dengan metode fast learner seperti Decision Tree, Naïve Bayes, ANN, dan SVM, teknik IBL dikenal sebagai lazy learner. Secara umum, metode K-NN terdiri dari dua langkah: menyimpan setiap pola latih saat pelatihan, dan memeriksa semua pola latih untuk menemukan sejumlah k pola terdekat saat melakukan klasifikasi sebuah pola. Nilai K yang diperoleh dari prosedur K-NN menawarkan akurasi generalisasi terbaik untuk data yang akan datang. Menemukan nilai K yang ideal adalah salah satu kesulitan dalam pelatihan. Bergantung pada jenis atribut data, nilai K dapat dihitung dengan menggunakan jarak, kemiripan, dan perbedaan. **27** Rumus berikut menjelaskan cara menghitung jarak Euclidean, yang digunakan untuk menentukan jarak antara dua titik. Gambar 2. 3 Rumus Perhitungan KNN 13

Langkah-langkah klasifikasi dapat diringkas sebagai berikut: 1) Memilih nilai k. 2) Menghitung rentang data antara testing dan tetangga terdekat nilai k pada data training. 3) Menentukan kelas yang memiliki jumlah tetangga terdekat terbanyak, dilakukan pemeriksaan dengan cara voting. **30** Kelas yang memiliki jumlah tetangga terdekat terbanyak akan dipilih sebagai hasil klasifikasi.

Performa metode kNN dalam klasifikasi biasanya dievaluasi menggunakan metrik akurasi. Metrik akurasi dapat dihitung dengan menggunakan rumus berikut. Gambar 2. 4 Rumus Akurasi KNN Ukuran kesederhanaan tertinggi adalah skor akurasi, yang mengukur jumlah prediksi benar dibandingkan dengan total kasus. Saat seluruh prediksi tepat, skor ini mencapai 1.0, tetapi jika semua prediksi salah, skornya menjadi 0. Sayangnya, skor akurasi ini memiliki kekurangan karena tidak memperhitungkan ketidakseimbangan antara false positive dan false negative. 2.1 **33** 11 Python Python

adalah salah satu bahasa komputer yang paling populer. Bahasa ini digunakan di banyak bidang, seperti pembelajaran mesin, pengembangan web, analisis data, dan pengembangan perangkat lunak secara umum. Guido van Rossum adalah penciptanya, dan pada awalnya dirilis pada tahun 1991. Python terkenal

karena sintaksnya yang sederhana dan mudah dipahami, serta menyediakan banyak pustaka dan modul untuk berbagai keperluan pengembangan perangkat lunak. Selain itu, Python memiliki komunitas pengguna yang besar dan aktif, menjadikannya salah satu bahasa pemrograman paling disukai di kalangan pengembang (Irfan Ardiansah & Ryan Hara Permana, 2023). Sebagai bahasa interpreter, Python tidak memerlukan kompilasi (mengubah kode sumber menjadi kode biner). Namun, Python mengubah kode sumber ke bentuk intermediate yang dapat dijalankan di berbagai platform komputer. Hal ini membuat Python lebih portabel. 11 Python termasuk dalam kategori bahasa pemrograman tingkat tinggi yang mendukung berbagai paradigma pemrograman seperti berorientasi objek, imperatif, dan fungsional. Salah satu keunggulan Python adalah sifatnya yang dinamis dengan manajemen memori otomatis (Syamsudin Manai, 2019).

2.2 Tinjauan Studi Penelitian ini didukung oleh berbagai referensi dari jurnal-jurnal terdahulu yang relevan dengan topik penelitian. Berikut adalah beberapa jurnal tersebut.

1. Jurnal pertama merupakan sebuah penelitian yang dilakukan oleh Adhitya Rahmat D 14 N, Karina Auliasari, Yosep Agus Pranoto. dengan judul 4 “IMPLEMENTASI METODE K-NEAREST NEIGHBOR (KNN) UNTUK SELEKSI CALON KARYAWAN BARU (Studi Kasus: BFI Finance Surabaya) 14 yang dimuat dalam Jurnal JATI (Jurnal Mahasiswa Teknik Informatika) pada tahun 2020. Jurnal ini membahas sebuah penelitian yang bertujuan untuk mengatasi 14 permasalahan dalam proses seleksi calon karyawan di BFI Finance Surabaya. Proses seleksi yang melibatkan berkas masuk, tes psikologi, interview, dan pemeriksaan kesehatan seringkali memakan waktu lama dan kurang efektif. Studi kasus dilakukan di BFI Finance Surabaya, di mana data calon karyawan baru dikumpulkan dan dianalisis. Hasil penelitian menunjukkan bahwa implementasi KNN dapat membantu dalam mengidentifikasi kandidat yang paling cocok dengan profil yang diinginkan oleh perusahaan. Proses seleksi menjadi lebih efisien dan akurat karena KNN dapat mempertimbangkan multiple atribut dan variabel dalam menentukan kemiripan antara calon karyawan dengan profil yang telah ada. Selain itu, jurnal ini juga mencakup evaluasi kinerja metode KNN dalam konteks seleksi

karyawan, seperti tingkat akurasi, sensitivitas, dan spesifisitas. Dengan menggunakan metode ini, BFI Finance Surabaya dapat mengoptimalkan proses rekrutmen dan seleksi karyawan baru, meningkatkan keberlanjutan sumber daya manusia, serta mengurangi risiko kesalahan dalam pemilihan calon karyawan. Implementasi KNN sebagai bagian dari sistem seleksi karyawan memberikan kontribusi positif terhadap efektivitas dan efisiensi proses rekrutmen perusahaan.  2.

Jurnal kedua merupakan sebuah penelitian yang dilakukan oleh Lusi Suryadi, Ngajiyanto, Novia Eka Pratiwi, Ferly Ardhy dan Pakartika Riswanto. dengan judul  3
“PENERAPAN DATA MINING PREDIKSI PENJUALAN MEBEL TERLARIS MENGGUNAKAN METODE K-NEAREST NEIGHBOR(K-NN) (STUDI KASUS: TOKO ZERITA MEUBEL)  4 yang dimuat dalam Jurnal JUSIM (Jurnal Sistem Informasi Musirawas) pada tahun 2022. Jurnal ini mengulas sebuah penelitian yang bertujuan untuk mengatasi masalah prediksi penjualan produk terlaris di Toko Zerita Meubel. Toko ini menjual berbagai jenis produk mebel dan, berdasarkan permintaan konsumen selama setahun terakhir, pemilik toko sudah bisa mengidentifikasi beberapa produk yang laris dan yang kurang laris. Penelitian ini didedikasikan untuk pengembangan sistem prediksi penjualan produk terlaris yang menggunakan teknik data mining dan metode k-nearest neighbour (K-NN), meskipun pada kenyataannya sistem yang ada saat ini masih bersifat manual dan menghasilkan data yang kurang akurat dan efisien. Di Toko Zerita Meubel, manajemen stok dan efektivitas strategi pemasaran dipengaruhi secara positif oleh data mining dengan metode K-NN. Sistem ini meningkatkan efisiensi operasional toko, mengurangi risiko kelebihan atau kekurangan stok, dan memberikan layanan yang lebih personal kepada pelanggan dengan menyediakan produk mebel sesuai preferensi mereka. Penelitian ini turut mengoptimalkan strategi bisnis dan meningkatkan keuntungan toko mebel tersebut. 3. Jurnal ketiga merupakan sebuah penelitian yang dibuat oleh Inna Alvi Nikmatun dan Indra Waspada. dengan judul “IMPLEMENTASI DATA MINING UNTUK KLASIFIKASI MASA STUDI MAHASISWA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR yang dimuat dalam Jurnal Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer pada tahun 2019.

Publikasi ini meneliti bagaimana masa studi mahasiswa diklasifikasikan dengan menggunakan teknik penggalian data. Penelitian ini bertujuan memahami dan memprediksi durasi studi mahasiswa berdasarkan atribut nilai mata kuliah mereka. Sebanyak 377 data siswa digunakan untuk proses pengambilan sampel ini; data ini mencakup 72 kualitas nilai mata kuliah dan 1 kelas target yang diwakili oleh periode studi, dan mereka berasal dari kelas yang lulus antara tahun 2007 dan 2011. Penambangan data adalah langkah yang dilakukan setelah penemuan pengetahuan dalam basis data (KDD). Untuk mengklasifikasikan masa studi, 15 algoritma K-Nearest Neighbour (K-NN) diterapkan. Cara terbaik untuk mengklasifikasikan masa studi didasarkan pada hasil percobaan yang memilih karakteristik dari semua kelas pilihan dengan tingkat keberhasilan 74,95%. Implikasi penting dalam pemantauan masa studi mahasiswa dimiliki oleh penelitian ini dan dapat membantu perguruan tinggi mengambil tindakan yang tepat untuk meningkatkan efisiensi studi. Langkah-langkah lebih efektif dapat diambil oleh institusi pendidikan untuk memastikan mahasiswa menyelesaikan studi dalam batas waktu yang ditetapkan dengan memahami faktor-faktor yang memengaruhi masa studi.

4. Jurnal keempat merupakan sebuah publikasi yang telah dilaksanakan oleh Aah Sumiah dan Nita Mirantika. dengan judul “Perbandingan Metode K-Nearest Neighbor dan Naive Bayes untuk Rekomendasi Penentuan Mahasiswa Penerima Beasiswa pada Universitas Kuningan yang dimuat dalam Jurnal Buffer Informatika pada tahun 2020. Publikasi ini membandingkan dan mengkontraskan algoritma K-Nearest Neighbour (KNN) dan Naive Bayes untuk merekomendasikan penerima beasiswa mahasiswa. Tujuan dari penelitian ini adalah untuk mengidentifikasi metode terbaik yang digunakan oleh Universitas Kuningan untuk memilih kandidat beasiswa. Bagian akademik dapat menggunakan sistem informasi yang dibuat dari data yang telah dianalisa untuk merekomendasikan beasiswa kepada pelamar dengan memanfaatkan Visual Basic.NET dan SQL Server. Penelitian ini merupakan pengembangan dari penelitian sebelumnya. Dalam studi ini, kedua algoritma dibandingkan untuk menilai efektivitasnya dalam klasifikasi data mahasiswa penerima beasiswa.

Pendekatan Naive Bayes mengungguli metode K-Nearest Neighbour, yang memiliki akurasi 40%, menurut hasil penelitian, dengan akurasi 70%. Penelitian ini memberikan wawasan berharga mengenai penerapan kedua algoritma tersebut dalam rekomendasi penerimaan beasiswa, yang dapat mendukung pengambilan keputusan di Universitas Kuningan.

5. Jurnal kelima merupakan sebuah penelitian yang dilakukan oleh Hozairi, Anwari dan Syariful Alim. dengan judul “IMPLEMENTASI ORANGE DATA MINING UNTUK KLASIFIKASI KELULUSAN MAHASISWA DENGAN MODEL K-NEAREST NEIGHBOR, DECISION TREE SERTA NAIVE BAYES yang dimuat dalam Jurnal Network Engineering Research Operation pada tahun 2023. Artikel ini membahas penggunaan pendekatan data mining untuk klasifikasi kelulusan siswa, khususnya dengan menggunakan alat Orange Data Mining. Pohon keputusan, Naive Bayes, dan K-Nearest Neighbor (K-NN) adalah tiga model kategorisasi utama, diintegrasikan oleh penelitian ini untuk meningkatkan akurasi dan keandalan prediksi kelulusan mahasiswa. Pola berdasarkan kemiripan dengan contoh yang ada diidentifikasi oleh metode K-NN, sedangkan aturan keputusan berbentuk pohon dihasilkan oleh Decision Tree. Teorema Bayes digunakan oleh model Naive Bayes untuk memperkirakan probabilitas kelulusan berdasarkan fitur-fitur yang diberikan. Penelitian ini dilakukan pada mahasiswa angkatan 2016 Program Studi Teknik Informatika Universitas Islam Madura. Penelitian ini bertujuan untuk menilai kekuatan prediksi dari tiga model - Decision Tree, Naive Bayes, dan K-Nearest Neighbor (K-NN) - dalam hal prediksi kelulusan mahasiswa. Model evaluasi Confusion Matrix dan ROC diimplementasikan selama prosedur pengujian data, yang dilakukan dengan menggunakan K-Fold Cross Validation (K=5). Hasil perbandingan antara ketiga model tersebut menunjukkan bahwa Naive Bayes lebih akurat (89% vs 77%) dibandingkan K-NN (77% vs 74%) dan Decision Tree 74% vs 74%. Penelitian ini menyarankan untuk menggunakan model 16 Naive Bayes untuk mengklasifikasikan tingkat kelulusan mahasiswa di Program Studi Teknik Informatika Universitas Islam Madura.

6. Jurnal keenam merupakan sebuah jurnal yang berjudul “Penerapan Data Mining untuk Memprediksi Tingkat Penjualan pada Produk Layanan Spesialis Kecantikan Menggunakan Metode K-Nearest Neighbor yang ditulis

oleh Richard Parulian Hutapea, Analisis komprehensif dilakukan oleh penulis untuk memprediksi tingkat penjualan produk layanan pada PT. Joanne Studio adalah studio kosmetik yang berfokus pada ekstensi bulu mata. Berbagai masalah dalam manajemen stok dihadapi oleh PT. Joanne Studio karena belum memiliki sistem yang dapat menghasilkan prediksi penjualan yang akurat. Penumpukan barang yang tidak terjual, kerusakan barang karena disimpan terlalu lama, atau bahkan kekurangan stok saat permintaan tinggi sering terjadi sebagai akibatnya. Perusahaan mengalami kerugian sebagai akibat dari semua ini. Penelitian ini bermaksud untuk menggunakan metode data mining dan algoritma K-Nearest Neighbor (KNN) untuk meramalkan tingkat penjualan untuk memecahkan masalah ini. Pengumpulan data penjualan selama enam bulan terakhir dilibatkan oleh penelitian ini yang kemudian dianalisis dengan teknik klasifikasi data mining menggunakan algoritma KNN. Mengolah data penjualan menjadi prediksi yang akurat adalah tujuan utama penerapan metode ini yang diharapkan dapat membantu perusahaan dalam perencanaan stok barang secara lebih efisien. Tingkat akurasi prediksi yang sangat tinggi, yaitu 94 persen, dicapai oleh para peneliti dengan menggunakan KNN. Bahwa model prediksi yang dibangun memiliki kemampuan yang baik dalam memperkirakan tingkat penjualan di masa depan ditunjukkan oleh skor akurasi ini sehingga dapat dijadikan sebagai dasar untuk pengambilan keputusan yang lebih akurat dalam manajemen stok. 7.

Penelitian yang dilakukan oleh Qurotul A'yuniyah dan Muhammad Reza pada tahun 2023 berjudul "Penerapan Algoritma K-Nearest Neighbor untuk Klasifikasi Jurusan Siswa di SMA Negeri 15 Pekanbaru bertujuan untuk mengatasi permasalahan yang dihadapi SMA Negeri 15 Pekanbaru dalam proses penjurusan siswa yang memakan banyak waktu dan tenaga ketika dilakukan secara manual. Algoritma K-Nearest Neighbor (KNN) dan alat RapidMiner digunakan oleh penelitian ini untuk mengklasifikasikan data siswa secara efisien. Hasil penelitian menunjukkan bahwa dengan parameter $k=3$, akurasi maksimal yang dicapai adalah 93.52%, rata-rata precision sebesar 88.14%, dan rata-rata recall mencapai 100.00%. Angka- angka ini menunjukkan bahwa algoritma KNN mampu

memberikan hasil klasifikasi yang sangat akurat dan andal untuk menentukan jurusan siswa dengan cepat dan tepat. Penggunaan KNN dalam klasifikasi jurusan siswa di SMA Negeri 15 Pekanbaru terbukti oleh kesimpulannya efektif dan efisien, menawarkan solusi praktis bagi sekolah untuk meningkatkan proses penjurusan siswa. Kontribusi penting dalam bidang Data Mining dan pendidikan diberikan oleh penelitian ini dengan menunjukkan bahwa teknik klasifikasi otomatis dapat meningkatkan akurasi dan efisiensi proses penjurusan, serta mendukung pengambilan keputusan yang lebih baik di lingkungan sekolah.

8. Penelitian yang dilakukan oleh Ma' ruf Aziz Muzani, M. Iqbal Abdullah Sukri, Syifa Nur Fauziah, Agus Fatkhurohman, dan Dhani Ariatmanto pada tahun 2021 dengan judul "Data Mining untuk Klasifikasi Produk Menggunakan Algoritma K-Nearest Neighbor pada Toko Online bertujuan untuk mengatasi tantangan dalam pencarian produk di e-marketplace yang semakin berkembang dan beragam, khususnya di platform Shopee. Investigasi ini menggunakan 17 algoritma K-Nearest Neighbour (KNN) untuk mengkategorikan produk dalam tiga kategori: peralatan rumah tangga, pakaian muslim, dan ponsel dan aksesoris. Data diperoleh melalui penggalan web. Preprocessing dilakukan terlebih dahulu dalam proses klasifikasi, yang kemudian diikuti dengan pembobotan kata dengan pendekatan TF-IDF dan mencari tahu seberapa mirip tulisan satu dengan yang lain menggunakan cosine similarity. Nilai k yang ideal ditemukan dengan membuat peringkat kemiripan kosinus. Setelah melakukan pengujian terhadap sembilan data dengan tiga nilai k yang berbeda, kami menemukan bahwa k=3 memberikan hasil terburuk dalam hal akurasi, presisi, dan recall. **26 Hasil yang lebih baik, dengan akurasi 88,89%, presisi 83,33%, dan recall 100%, diperoleh dengan nilai k=5 dan k=7.**

Temuan ini menunjukkan bahwa produk di e-marketplace dapat diklasifikasikan secara efektif menggunakan algoritma KNN, memudahkan pengguna dalam mencari produk yang diinginkan dengan lebih cepat dan akurat. Kontribusi penting dalam penerapan data mining diberikan oleh penelitian ini untuk meningkatkan pengalaman berbelanja online dengan menyediakan metode yang efisien untuk pengelompokan produk berdasarkan

kemiripan atribut. 18 BAB III METODE PENELITIAN 3.1 Alasan Penggunaan Metode Penggunaan metode klasifikasi dengan algoritma K-Nearest Neighbor (K-NN) didasarkan pada efektivitas dan efisiensi algoritma ini dalam menangani data besar dan kompleks yang sering ditemui dalam penerimaan mahasiswa baru. K-NN tidak hanya efisien dalam mengklasifikasikan dan menganalisis data, tetapi juga memiliki keunggulan dalam memprediksi preferensi calon mahasiswa, yang sangat penting untuk pengembangan strategi pemasaran dan program studi di universitas. Fleksibilitas K-NN dalam pemilihan fitur dan penyesuaian parameter memungkinkan penelitian ini untuk disesuaikan dengan kebutuhan data dan pertanyaan penelitian. Selain itu, kemudahan interpretasi hasil yang ditawarkan oleh K-NN sangat penting untuk memastikan bahwa temuan penelitian dapat dengan mudah dipahami dan diterapkan oleh pihak universitas dalam pengambilan keputusan. Sifat K-NN yang sederhana namun efektif ini cocok untuk bekerja dengan data yang memiliki banyak variabel, seperti yang umum ditemui dalam data penerimaan mahasiswa, memberikan penjelasan yang lebih menyeluruh dan tepat mengenai elemen-elemen yang mempengaruhi keputusan penerimaan siswa. Menurut (P.W Rahayu, dkk. 2024) menjelaskan bahwa tidak ada alasan utama yang mutlak untuk pemilihan algoritma tertentu dalam penelitian ini, pemilihan algoritma sebaiknya disesuaikan dengan karakteristik kasus yang ditemukan. Dalam kasus ini, penggunaan K-NN lebih cocok karena kesederhanaan dan efektivitasnya untuk data dengan banyak variabel yang tidak saling berkaitan. Meskipun K-NN memiliki kelebihan dalam menangani jenis data ini, penting untuk melakukan evaluasi kembali untuk memastikan bahwa algoritma yang dipilih benar-benar yang terbaik. Evaluasi ini melibatkan pengujian dan perbandingan dengan algoritma lain untuk memastikan akurasi dan efektivitas yang optimal dalam konteks data penerimaan mahasiswa baru. Dengan demikian, peneliti dapat memastikan bahwa metode yang digunakan benar-benar memberikan hasil yang andal dan berguna untuk pengambilan keputusan di universitas. 3.2 Lokasi Penelitian Penelitian ini difokuskan pada Universitas XYZ, sebuah institusi pendidikan tinggi

terletak di salah satu kota terbesar di Indonesia. Universitas ini dipilih karena representatifnya dalam hal keragaman program studi dan populasi mahasiswa. Universitas XYZ dikenal karena sistem administrasinya yang terstruktur dengan baik, terutama dalam pengelolaan data penerimaan mahasiswa.

32 Ini memungkinkan penelitian untuk mengakses data yang dibutuhkan dengan efektif dan efisien. Dengan populasi mahasiswa yang beragam dan meluasnya program studi yang ditawarkan, Universitas XYZ menjadi lokasi yang ideal untuk menganalisis bagaimana minat calon mahasiswa terhadap berbagai program studi berubah seiring waktu dan faktor apa saja yang mempengaruhinya. 3.3 Instrumen Penelitian Dua komponen utama termasuk dalam instrumen penelitian yang digunakan. Pertama, Python, sebuah bahasa pemrograman tingkat tinggi, dijadikan alat analisis data utama karena kemampuannya yang luas dalam pemrograman dan analisis data. Kedua, dataset historis penerimaan mahasiswa baru di Universitas XYZ, yang mengandung informasi seperti data demografis dan pilihan program studi calon mahasiswa. Dataset ini akan diolah dengan Python untuk mengidentifikasi pola 19 yang relevan, memberikan wawasan yang penting untuk memahami minat dan pilihan calon mahasiswa. 3.4 Sampel atau Sumber Data Data untuk penelitian ini berasal dari divisi marketing Universitas XYZ. Sebelum mereka melakukan registrasi secara resmi, data yang digunakan adalah data lead atau target pendaftar.. Data ini dianggap memberikan pemahaman yang berharga tentang calon mahasiswa yang telah menunjukkan minat atau niat untuk bergabung dengan universitas sebelum proses pendaftaran formal dimulai. Variabel-variabel yang diambil, seperti jenis kelamin, jenis sekolah, asal sekolah, jenis seleksi, dan pilihan program studi, dipilih dengan hati-hati karena dianggap memberikan wawasan yang penting tentang preferensi calon mahasiswa dalam memilih program studi. 3.5 Teknik Pengumpulan Data Teknik pengumpulan data dimaksudkan untuk menjamin kelengkapan, keakuratan, dan relevansi informasi yang digunakan. Langkah pertama dalam proses ini adalah memperoleh akses ke database Universitas XYZ, yang melibatkan prosedur izin yang ketat

untuk mematuhi standar etika dan privasi. Setelah akses terjamin, data yang relevan untuk penelitian akan diekstrak, yaitu Data Penerimaan Mahasiswa Baru. Proses ekstraksi diikuti dengan langkah pembersihan dan pemrosesan data untuk menghilangkan kesalahan dan ketidakakuratan, memastikan bahwa data yang dianalisis berkualitas tinggi dan dapat diandalkan. Data ini kemudian dikodekan dan dikategorikan sesuai dengan kebutuhan analisis, seperti demografi dan pilihan program studi.

3.6 Teknik Analisis Data

Untuk memahami pola dan hubungan antar faktor dalam kumpulan data penerimaan siswa baru, Anda perlu menganalisis data tersebut. **18** Teknik analisis data digunakan dalam penelitian ini untuk menemukan informasi yang relevan dan membantu dalam pengambilan keputusan. Proses klasifikasi menggunakan algoritma K-Nearest Neighbors (KNN) melibatkan beberapa langkah utama untuk memastikan model yang dihasilkan akurat dan andal. Berikut adalah langkah-langkah detail dari proses tersebut:

1. Mengumpulkan Data Langkah pertama dalam setiap proyek klasifikasi adalah mengumpulkan dataset yang relevan. Dataset ini harus mencakup fitur-fitur yang akan digunakan untuk prediksi dan label kelas yang akan diprediksi. Dataset dapat berasal dari sumber publik atau dikumpulkan sendiri melalui eksperimen atau survei.
2. Pra-pemrosesan Data Data mentah seringkali membutuhkan beberapa tahap pra-pemrosesan sebelum bisa digunakan untuk membangun model KNN.
 - Normalisasi/Penskalaan Data: Melakukan normalisasi atau penskalaan pada data karena KNN sangat sensitif terhadap skala fitur. Ini memastikan bahwa semua fitur memiliki kontribusi yang seimbang dalam perhitungan jarak.
 - 20 - Menghandle Missing Values: Menangani nilai yang hilang dalam dataset, misalnya dengan menggantinya menggunakan metode imputasi seperti mean atau median.
3. Memilih Nilai K Untuk menentukan nilai K yang paling ideal, sejumlah nilai K yang berbeda dieksplorasi, dan validasi silang digunakan. Overfitting dapat terjadi pada nilai K yang kecil, sementara underfitting dapat terjadi pada nilai K yang besar. Dalam KNN, nilai K- jumlah tetangga terdekat-harus dipilih dengan hati-hati.
4. Mendefinisikan Data Latih dan Data Uji Informasi dibagi

menjadi dua bagian: data uji dan data pelatihan, sebelum model dibangun. Biasanya, pembagian ini dilakukan dengan perbandingan 70:30 atau 80:20.

28 Data yang belum pernah dilihat model selama pelatihan digunakan untuk memastikan evaluasi yang objektif. 24 5. Menghitung Jarak Dalam langkah ini, hitung jarak antara setiap titik data uji dengan semua titik data latih. Jarak

Euclidean sering menjadi pilihan karena kesederhanaannya: *Jarak* 

Euclidean = $\sqrt{(x_1 - y_1)^2 + \dots + (x_n - y_n)^2}$ Kete

rangan : - $X = (x_1, x_2, \dots, x_n)$ adalah vektor fitur dar

i titik data pertama. - $Y = (y_1, y_2, \dots, y_n)$ adalah

vektor fitur dari titik data kedua. - x_i dan y_i adalah nilai

fitur ke- i dari titik data X dan Y masing-masing. - $\sum (x$

$- y)^2$ adalah jumlah dari kuadrat perbedaan antara nilai-n

ilai fitur yang sesuai dari kedua titik data. 6. Menentukan K

Tetangga Terdekat K tetangga terdekat dari titik data uji diidentifikasi

setelah jarak dihitung. Lokasi- lokasi dalam set pelatihan yang paling

dekat dengan set pengujian disebut tetangga. 7. Voting Setelah dilakukan

voting untuk menentukan kelas dari titik data uji, prediksi untuk titik

data uji tersebut adalah kelas yang paling sering terlihat antara K

tetangga terdekat. 21 8. Evaluasi Model Evaluasi performa model adalah

langkah terakhir. Mengukur akurasi, presisi, recall, dan F1-score dilakukan

dengan data uji. Ini memberikan gambaran yang jelas mengenai kinerja

model. Teknik validasi silang digunakan untuk memastikan konsistensi hasil.

22 BAB IV HASIL DAN PEMBAHASAN 4.1 Data Understanding Dataset yang digunakan dalam penelitian ini diperoleh langsung dari bagian pemasaran

Universitas XYZ. Data tersebut mencakup periode dari tahun 2020 hingga

2023, memberikan peneliti gambaran menyeluruh tentang peminatan program

studi selama beberapa tahun terakhir. Dengan jumlah total record sebanyak

13.404, setiap entri data mewakili informasi lengkap tentang calon

mahasiswa termasuk berbagai fitur. Rentang waktu yang dipilih tidak hanya

memberikan penelitian akses terhadap data yang relevan dalam kurun waktu

yang cukup luas, tetapi juga memungkinkan analisis yang lebih mendalam

terhadap perubahan yang terjadi dari tahun ke tahun. Dengan pemahaman yang mendalam tentang karakteristik dataset, peneliti dapat mengarahkan penelitian dengan lebih tepat sasaran, serta merencanakan strategi analisis yang efektif untuk menemukan wawasan dan pengetahuan yang bernilai. Penelitian ini akan menghasilkan beberapa output utama. Nilai akurasi dari K-Nearest Neighbor adalah salah satunya (KNN) yang dihitung untuk mengevaluasi kinerja model dalam memprediksi peminatan program studi. Akurasi ini akan memberikan indikasi seberapa baik model dapat mengklasifikasikan data baru berdasarkan data historis. Selain itu, grafik dan persentase pertumbuhan juga akan dihasilkan untuk menggambarkan distribusi peminatan program studi dari tahun ke tahun, serta untuk menunjukkan tren dan perubahan yang terjadi. Persentase pertumbuhan ini akan memberikan pemahaman lebih lanjut tentang dinamika peminatan program studi di Universitas XYZ selama periode yang diteliti. Diharapkan bahwa temuan ini akan memberikan pemahaman yang komprehensif tentang tren spesialisasi program studi dan membantu Universitas XYZ dalam merumuskan strategi pemasaran dan perencanaan akademik yang lebih efektif. Tabel 4.

1 Data Understanding 4.2 Data Selection Tahap ini melibatkan pemilihan variabel-variabel yang relevan untuk membangun model prediksi peminatan program studi dengan menggunakan metode K-Nearest Neighbour (KNN). Proses pemilihan variabel ini bertujuan untuk menangkap karakteristik yang signifikan dari calon mahasiswa baru yang dapat mempengaruhi pemilihan program studi mereka. Variabel-variabel yang dipilih dalam proses ini meliputi berbagai aspek yang mencakup latar belakang pendidikan, jalur penerimaan, dan preferensi individu calon mahasiswa. Variabel-variabel yang dipilih adalah sebagai berikut:

1. Tahun Penerimaan Tahun akademik saat calon mahasiswa mengisi data kuisisioner peminatan program studi. Variabel ini penting untuk melihat peminatan program studi dari tahun ke tahun dan memahami bagaimana perubahan dalam kebijakan penerimaan atau kondisi eksternal mempengaruhi pilihan calon mahasiswa.
2. Jenis Kelamin 23 Jenis kelamin calon mahasiswa. Variabel ini digunakan untuk mengidentifikasi pola

peminatan berdasarkan gender, yang dapat memberikan wawasan tentang kecenderungan pemilihan program studi antara calon mahasiswa laki-laki dan perempuan. 3. Jenis Sekolah Jenis sekolah asal calon mahasiswa. Variabel ini membantu dalam memahami pengaruh latar belakang pendidikan terhadap pemilihan program studi. 4. Asal Sekolah Lokasi sekolah asal calon mahasiswa. Variabel ini berguna untuk mengidentifikasi pengaruh geografis dan institusional terhadap peminatan program studi. Misalnya, sekolah-sekolah dari daerah tertentu mungkin memiliki kecenderungan untuk memilih program studi tertentu di universitas. 5. Jalur Seleksi Jalur penerimaan calon mahasiswa. Variabel ini penting untuk melihat perbedaan peminatan berdasarkan jalur seleksi yang diikuti oleh calon mahasiswa. Setiap jalur seleksi mungkin memiliki karakteristik atau standar yang berbeda yang dapat mempengaruhi pilihan program studi. 6. Jurusan Sekolah Jurusan yang diambil calon mahasiswa saat di sekolah jenjang menengah atas. Variabel ini digunakan untuk memahami bagaimana latar belakang jurusan di sekolah menengah mempengaruhi pemilihan program studi di universitas. Misalnya, calon mahasiswa dari jurusan IPA mungkin lebih cenderung memilih program studi sains atau teknik. 7. Program Studi yang Dipilih Program studi pilihan calon mahasiswa pada saat mendaftar. Variabel ini adalah variabel utama yang menjadi target prediksi dalam model KNN. Informasi ini mencerminkan preferensi awal calon mahasiswa dalam memilih program studi di universitas. Selain variabel-variabel tersebut, proses ini juga menambahkan variabel label sebagai acuan untuk mengelompokkan program studi berdasarkan tingkat peminatan. Variabel label ini dikategorikan menjadi dua kelompok utama Minat dan Kurang Diminati. Label Minat mencakup program studi yang dianggap lebih diminati oleh calon mahasiswa baru. Program studi yang termasuk dalam kategori ini adalah Ilmu Komunikasi, Manajemen, dan Psikologi. Ketiga program studi ini sering kali menjadi pilihan utama karena salah satu faktor utamanya adalah popularitas di kalangan calon mahasiswa. Di sisi lain, Label Kurang Diminati mencakup program studi yang dianggap kurang diminati 12 oleh calon mahasiswa baru. Program

studi dalam kategori ini adalah Sistem Informasi, Informatika, Teknik Sipil, Arsitektur, Akuntansi, Desain Komunikasi Visual, dan Desain Produk.

Meskipun program studi ini memiliki prospek yang baik dan relevansi industri yang kuat, mereka mungkin kurang diminati oleh sebagian calon mahasiswa baru karena berbagai alasan seperti persepsi tentang tingkat kesulitan, kurangnya informasi, atau minat pribadi yang lebih rendah dalam bidang-bidang tersebut. Hasil dari proses seleksi data ditampilkan pada Gambar 4.1 di bawah ini. 24 Gambar 4. 1 Data Selection 4.3

Data Cleaning Prosedur ini bertujuan untuk menjamin data yang bebas dari kesalahan untuk penyelidikan. atau inkonsistensi yang dapat mempengaruhi hasil analisis. Proses pembersihan data ini merupakan langkah penting untuk meningkatkan kualitas data dan memastikan akurasi prediksi peminatan program studi calon mahasiswa baru. Langkah-langkah utama dalam pembersihan data meliputi menghilangkan missing values dan menghapus duplikat data. 1. Menghilangkan Missing Values Menghilangkan missing values adalah langkah kritis dalam proses data cleaning. Missing values dapat muncul dalam berbagai bentuk, seperti sel kosong atau entri dengan nilai "null" atau "N/A". Identifikasi missing values dilakukan dengan menggunakan fungsi-fungsi statistik yang mendeteksi sel-sel kosong atau tidak lengkap dalam dataset. Setelah mengidentifikasi missing values, penyebabnya dianalisis untuk menentukan pendekatan yang tepat dalam menanganinya. Dalam konteks penelitian ini, missing values dapat disebabkan oleh berbagai faktor, seperti kesalahan saat pengumpulan data atau entri yang tidak diisi oleh calon mahasiswa. Beberapa strategi dapat digunakan untuk menangani missing values, termasuk isi nilai yang hilang atau gunakan alternatif yang sesuai seperti rata-rata atau median untuk catatan yang tidak memilikinya.. Namun, dalam penelitian ini, diputuskan untuk menghapus record yang memiliki missing values guna memastikan data yang digunakan sepenuhnya lengkap dan akurat. Langkah ini diambil karena jumlah record yang memiliki missing values relatif kecil dan tidak signifikan terhadap ukuran keseluruhan dataset. Proses penghapusan missing values dimulai

dengan mengidentifikasi dan mencatat semua record yang tidak lengkap. Setelah itu, record tersebut dihapus dari dataset. Dari total 13.404 record data calon mahasiswa baru, proses pembersihan ini mengakibatkan pengurangan jumlah data menjadi 12.008 record. Dengan menghilangkan 1.396 record yang memiliki missing values, dataset yang tersisa lebih lengkap dan konsisten, yang akan meningkatkan akurasi dan keandalan model prediksi yang dikembangkan.

2. Menghapus Data Duplikat

Menghapus data duplikat adalah langkah krusial untuk memastikan bahwa setiap entri dalam dataset adalah unik dan valid. Data duplikat adalah entri yang identik yang muncul lebih dari sekali dalam dataset dengan informasi yang sama. Identifikasi duplikat dilakukan dengan menggunakan fungsi-fungsi khusus dalam perangkat lunak pengolahan data, seperti `drop_duplicates` di Python atau fitur serupa di perangkat lunak lain. Setelah mengidentifikasi data duplikat, penyebab kemunculan duplikat dianalisis untuk memahami mengapa duplikasi terjadi. Penyebab umum termasuk kesalahan saat pengumpulan data, kesalahan input manual, atau duplikasi data saat integrasi dari berbagai sumber. Memahami penyebab ini membantu dalam mengembangkan prosedur untuk mencegah terjadinya duplikasi di masa mendatang.

25 Setelah duplikat diidentifikasi, langkah selanjutnya adalah menghapus record yang duplikat sehingga hanya satu entri unik yang tersisa. Dalam penelitian ini, penghapusan duplikat dilakukan secara otomatis menggunakan fungsi perangkat lunak yang sesuai. Proses ini mengakibatkan pengurangan jumlah data lebih lanjut. Dari total 12.008 record yang telah dibersihkan dari missing values, identifikasi dan penghapusan duplikat mengurangi jumlah record menjadi 6.109. Berikut merupakan jumlah data sebelum dan sesudah proses cleaning dilakukan.

Gambar 4. 2 Data Before Cleaning Gambar 4. 3 Data After Cleaning

4.4 Data Transformation

Pada tahap berikutnya, proses transformasi data dilakukan untuk mempersiapkan data agar dapat digunakan dalam model prediksi peminatan program studi dengan metoda K-Nearest Neighbor (KNN). Proses transformasi data ini sangat penting karena metode KNN bekerja dengan data numerik, sehingga semua variabel dalam dataset harus diubah

ke dalam format numerik. Pada penelitian ini, semua variabel yang digunakan merupakan tipe data objek kategorikal yang perlu diubah menjadi numerik. Untuk mengubah variabel kategorikal menjadi numerik, digunakan teknik encoding. Teknik yang digunakan adalah Label Encoding. Teknik ini memberikan label numerik unik untuk setiap kategori dalam variabel. Misalnya, untuk variabel Jenis Kelamin, kategori "Laki-laki" dapat diubah menjadi 0 dan kategori "Perempuan" dapat diubah menjadi 1. Label Encoding sederhana dan cocok untuk variabel dengan urutan kategori yang inheren. Gambar 4. 4 Data After Transformation Dengan melakukan transformasi data ini, dataset yang awalnya terdiri dari variabel-variabel kategorikal kini telah diubah menjadi bentuk numerik yang dapat diproses oleh model KNN. Transformasi data memastikan bahwa informasi yang terkandung dalam variabel kategorikal tetap terwakili dengan benar dalam bentuk numerik. Proses ini penting untuk meningkatkan akurasi dan efisiensi model prediksi, karena data numerik diperlukan oleh metode KNN untuk menghitung jarak antara titik data.

4.5 Penerapan Metode KNN

Metode K-Nearest Neighbour (KNN) diimplementasikan pada tahap ini untuk mengantisipasi peminatan calon mahasiswa baru pada program studi di Universitas XYZ. Metode KNN adalah algoritma klasifikasi yang digunakan untuk mengelompokkan data berdasarkan kedekatan atau kemiripan antara titik data. Proses penerapan KNN melibatkan beberapa langkah penting yang memastikan model prediksi bekerja dengan optimal dan menghasilkan hasil yang akurat.

4.5.1 Menentukan Variabel X dan Y

Langkah pertama dalam penggunaan metode K-Nearest Neighbor (KNN) adalah menentukan variabel input (X) dan variabel output (Y) setelah data penerimaan mahasiswa baru ditransformasi. Variabel-variabel yang digunakan untuk memprediksi peminatan program studi calon mahasiswa baru disebut variabel X. Variabel-variabel tersebut mencakup Tahun Penerimaan, Jenis Kelamin, Jenis Sekolah, Asal Sekolah, Jalur Seleksi, Jurusan Sekolah, dan Program Studi yang Dipilih. Setiap variabel memiliki peran penting dalam membentuk profil calon mahasiswa, yang pada gilirannya akan digunakan oleh model KNN untuk

melakukan prediksi. Langkah selanjutnya adalah menetapkan variabel Y. Variabel Y adalah label atau kategori yang ingin diprediksi oleh model, yaitu minat calon mahasiswa terhadap program studi tertentu. Penentuan variabel Y sangat penting karena model KNN akan belajar dari data historis untuk membuat prediksi yang akurat tentang minat program studi berdasarkan variabel X. Gambar 4. 5 Menentukan Variabel X dan Y

4.5.2 Mengatasi Imbalance Data Dengan Smote

Ketidakseimbangan data adalah masalah umum dalam machine learning yang terjadi ketika jumlah sampel dalam kategori yang berbeda sangat tidak merata. Dalam penelitian ini, terdapat ketidakseimbangan antara jumlah calon mahasiswa yang memilih program studi dalam kategori "Minat" dan "Kurang Diminati". Ketidakseimbangan ini dapat menyebabkan model K-Nearest Neighbor (KNN) menjadi bias terhadap kategori mayoritas, sehingga mengurangi akurasi prediksi **29** untuk kategori minoritas. Teknik Synthetic Minority Over Sampling Technique (SMOTE) digunakan untuk mengatasi masalah ini. SMOTE, sebuah metode oversampling, menciptakan sampel sintesis baru dari data minoritas yang ada. Dengan demikian, SMOTE membantu dalam menyeimbangkan distribusi data tanpa menambahkan sampel yang sama persis, yang dapat menyebabkan overfitting. Gambar 4. 6 Oversampling with SMOTE Gambar 4. 7 Variabel Minat setelah di SMOTE

4.5.3 Scaling Dataset

Menskalakan dataset adalah tahap selanjutnya dalam menggunakan pendekatan K-Nearest Neighbor (KNN) setelah ketidakseimbangan data dikoreksi menggunakan SMOTE. Scaling sangat penting dalam algoritma KNN karena metode ini menggunakan jarak Euclidean untuk menghitung kedekatan antar titik data. Perbedaan skala antara fitur-fitur dapat mempengaruhi perhitungan jarak ini, sehingga mempengaruhi kinerja model. **27** Untuk melakukan scaling, digunakan Standard Scaler. Dengan scaling, fitur-fitur yang memiliki rentang nilai yang sangat berbeda dapat disesuaikan sehingga semuanya berada dalam skala yang sama, memungkinkan model untuk memproses data secara lebih efektif dan akurat. Proses scaling dengan Standard Scaler melibatkan dua langkah utama, yaitu fitting dan transforming. Pertama, scaler di-fit pada data training untuk

menghitung mean dan standard deviation dari setiap fitur. Kedua, scaler tersebut digunakan untuk mentransformasi data training dan data testing berdasarkan mean dan standard deviation yang telah dihitung. Data yang dilakukan scaling pada data training dan data testing dipastikan konsisten oleh langkah ini, yang penting untuk menjaga integritas model prediksi. Gambar 4.8 Scalling Dataset

4.5.4 Menentukan Data Training dan Testing

Tahap berikutnya adalah membagi dataset menjadi dua, membuat set data yang terpisah untuk pelatihan dan pengujian. Pembagian ini dilakukan untuk mengevaluasi kinerja model secara obyektif. Parameter `test_size = 0.3` dan `random_state = 42` digunakan dalam penelitian ini untuk memastikan bahwa pembagian tersebut konsisten dan dapat direproduksi. Parameter `test_size=0.3` menunjukkan bahwa 30% dari total dataset dialokasikan sebagai data testing, sementara 70% sisanya digunakan sebagai data training. Ini memastikan bahwa data testing cukup besar untuk memberikan evaluasi yang representatif terhadap kinerja model, sementara data training tetap cukup besar untuk memungkinkan model belajar dengan baik dari pola yang ada. Penggunaan `random_state=42` memastikan bahwa pembagian dataset dilakukan secara acak namun konsisten setiap kali proses ini dijalankan, sehingga hasil evaluasi dapat dibandingkan secara adil dan konsisten. Gambar 4.9 Data Training and Test

4.5.5 Menentukan Nilai K

Kinerja model sangat dipengaruhi oleh pemilihan nilai k yang tepat. Jika k ditetapkan terlalu kecil, model akan menjadi terlalu sensitif terhadap noise; sebaliknya, jika k ditetapkan terlalu besar, model akan mengabaikan informasi yang signifikan dalam data. Untuk menentukan nilai k terbaik, digunakan grid search untuk menguji beberapa nilai k yang berbeda. Dalam penelitian ini, diuji nilai k yaitu 3, 5, 7, dan 9. Memilih nilai ganjil membantu menghindari masalah tie (seri) dalam pemilihan mayoritas di KNN. Dalam KNN, mayoritas label dari k tetangga terdekat menentukan hasil prediksi. Jika k bernilai genap, ada kemungkinan terjadi seri, di mana jumlah tetangga dari dua atau lebih kelas yang berbeda sama, sehingga sulit untuk menentukan kelas mayoritas.

2.8 Nilai k yang lebih

kecil seperti 3 membuat model lebih fokus pada data yang sangat dekat dengan titik yang diprediksi, sehingga model bisa menangkap detail yang lebih spesifik. Namun, nilai k yang kecil juga membuat model lebih mudah terpengaruh oleh data yang tidak biasa atau salah. Model menjadi lebih stabil dan kurang terpengaruh oleh data yang tidak biasa karena penggunaan nilai k yang lebih tinggi, seperti 9, meskipun ini bisa menyebabkan hilangnya beberapa rincian penting karena mempertimbangkan terlalu banyak data yang mungkin kurang relevan. Setelah melakukan grid search, didapatkan bahwa nilai k terbaik adalah 3. Pemilihan $k=3$ menunjukkan bahwa model memberikan performa terbaik dalam hal akurasi prediksi peminatan program studi calon mahasiswa baru ketika mempertimbangkan tiga tetangga terdekat. Nilai k ini memberikan keseimbangan optimal antara sensitivitas terhadap data dan kemampuan generalisasi, sehingga memberikan hasil prediksi yang paling akurat. Langkah penting dalam mengembangkan model KNN adalah menentukan nilai k yang sesuai, memastikan bahwa hasil yang akurat dan berguna dapat diberikan oleh model bagi Universitas XYZ untuk memahami dan memprediksi minat program studi calon mahasiswa baru. Gambar 4. 10 Menentukan Nilai k

4.5.6 Pelatihan Model KNN dan Classification Report

Dengan menetapkan nilai k , data pelatihan digunakan untuk mengajari model KNN. Proses pelatihan ini membantu model dalam memahami pola dan hubungan antara fitur input dan label minat. Sesudah pelatihan, model diuji dengan data uji untuk menilai kinerjanya. Mengevaluasi model dilakukan dengan membuat laporan klasifikasi yang mencakup berbagai metrik evaluasi seperti precision, recall, f1-score, dan akurasi. Laporan klasifikasi ini menyediakan gambaran lengkap tentang performa model dalam memprediksi setiap kategori minat. Berdasarkan hasil evaluasi, model KNN yang dilatih dengan nilai k terbaik memberikan akurasi sebesar 88%. Ini berarti model mampu memprediksi peminatan program studi calon mahasiswa baru dengan tingkat ketepatan yang tinggi. Akurasi yang dicapai menunjukkan bahwa model KNN yang dibangun efektif dalam mengenali pola-pola

penting dalam data dan memberikan prediksi yang andal. Gambar 4. 11 Akurasi Skor 4.6 Data Visualization Visualisasi data adalah alat yang berguna untuk analisis dan pemahaman data. Visualisasi data digunakan untuk memberikan gambaran yang jelas tentang jumlah peminat program studi dan pertumbuhan peminat program studi di Universitas XYZ dari tahun 2020 hingga 2023. Dengan visualisasi data, informasi yang terdapat dalam data dapat diungkap dengan lebih baik, 29 memungkinkan pembuatan keputusan yang lebih tepat dan pemahaman yang mendalam tentang pola dan tren yang muncul difasilitasi oleh visualisasi data. Gambar 4. 12

Perbandingan Data Calon Mahasiswa Grafik yang ditampilkan menunjukkan perbandingan Banyak calon mahasiswa baru yang mengajukan pendaftaran ke Universitas XYZ dari tahun 2020 hingga 2023. Berikut adalah penjelasan mengenai data dalam grafik tersebut: 1. Pada tahun 2020, jumlah calon mahasiswa yang mendaftar adalah yang paling sedikit dibandingkan tahun-tahun lainnya, dengan total sekitar 1.000 calon mahasiswa. Ini mungkin disebabkan oleh berbagai faktor, termasuk kemungkinan dampak awal pandemi COVID-19 yang menyebabkan ketidakpastian dalam rencana pendidikan banyak calon mahasiswa. 2. Jumlah calon mahasiswa meningkat signifikan pada tahun 2021, dengan total mencapai sekitar 2.500 calon mahasiswa. Peningkatan ini menunjukkan adanya pemulihan minat dalam melanjutkan pendidikan tinggi setelah periode ketidakpastian pada tahun sebelumnya. 3. Tren peningkatan jumlah calon mahasiswa berlanjut pada tahun 2022, dengan total mencapai sekitar 3.500 calon mahasiswa. Pertumbuhan ini dapat mencerminkan stabilisasi situasi pandemi dan adaptasi masyarakat terhadap kondisi normal baru, termasuk dalam sektor pendidikan. 4. Tahun 2023 menunjukkan puncak jumlah pendaftaran, dengan total sekitar 5.000 calon mahasiswa. Angka ini mencerminkan pertumbuhan yang paling tinggi dalam periode yang diteliti, menunjukkan kemungkinan keberhasilan strategi pemasaran dan promosi Universitas XYZ, serta meningkatnya calon mahasiswa tertarik untuk meneruskan studi mereka di universitas tersebut. Secara keseluruhan, grafik tersebut menggambarkan tren peningkatan jumlah calon

mahasiswa baru dari tahun ke tahun. Pertumbuhan yang konsisten setiap tahunnya menunjukkan adanya peningkatan kepercayaan dan minat calon mahasiswa terhadap Universitas XYZ. Faktor-faktor seperti adaptasi terhadap situasi pandemi, peningkatan kualitas promosi universitas, serta berbagai program dan fasilitas yang ditawarkan oleh universitas dapat berkontribusi terhadap tren positif ini. Selanjutnya, dipaparkan grafik jumlah peminat program studi di Universitas XYZ dari tahun 2020 hingga 2023. Visualisasi data akan membantu kita memahami tren minat mahasiswa baru terhadap berbagai program studi.

Gambar 4. 13 Jumlah Peminat Tahun 2020 Pada tahun 2020, data menunjukkan bahwa program studi dengan jumlah peminat tertinggi adalah Ilmu Komunikasi, yang memiliki jumlah peminat tertinggi. Posisi kedua ditempati oleh program studi Manajemen, sementara program studi Psikologi berada di posisi ketiga dalam hal jumlah peminat.

Gambar 4. 14 Jumlah Peminat Tahun 2021 Pada tahun 2021, data menunjukkan perubahan dalam jumlah peminat program studi. Program studi dengan jumlah peminat tertinggi adalah Manajemen. Posisi kedua ditempati oleh Ilmu Komunikasi, sementara Psikologi tetap berada di posisi ketiga dalam hal jumlah peminat.

Gambar 4. 15 Jumlah Peminat Tahun 2022 Pada tahun 2022, program studi Manajemen memiliki jumlah peminat tertinggi, diikuti oleh Ilmu Komunikasi. Kedua program studi ini hampir menyentuh 700 peminat. Psikologi tetap berada di posisi ketiga dalam hal jumlah peminat.

Gambar 4. 16 Jumlah Peminat Tahun 2023 Pada tahun 2023, terjadi perubahan signifikan dalam jumlah peminat program studi. Ilmu Komunikasi kembali menduduki peringkat pertama dengan hampir menyentuh 1000 peminat. Psikologi menempati posisi kedua, diikuti oleh Manajemen, keduanya juga hampir mencapai jumlah peminat yang sama. Sehingga berikut merupakan grafik pertumbuhan selama 2020 hingga 2023 secara keseluruhan yang ditampilkan pada Gambar 4.17 Data Peminat Keseluruhan.

Gambar 4. 17 Jumlah Peminat Keseluruhan Grafik tersebut menggambarkan peningkatan jumlah calon mahasiswa baru Universitas XYZ pada berbagai program studi **17** dari tahun 2020 hingga 2023. Sumbu horizontal

(x) digunakan untuk menunjukkan tahun, sedangkan sumbu vertikal (y) menunjukkan jumlah mahasiswa baru (count). Berdasarkan grafik, beberapa program studi menunjukkan tren peningkatan yang stabil dan signifikan dari tahun ke tahun. Misalnya, Akuntansi dan Informatika menunjukkan peningkatan yang konsisten, sementara Desain Produk mengalami peningkatan yang sangat signifikan, terutama pada tahun 2021 dan 2023. Di sisi lain, program studi seperti Arsitektur dan Manajemen menunjukkan peningkatan yang lebih moderat. Grafik ini memberikan pemahaman yang jelas tentang preferensi mahasiswa terhadap berbagai program studi serta perubahan tren dari masa ke masa yang bisa digunakan untuk analisis lebih lanjut terkait peminatan dan strategi penerimaan mahasiswa baru di Universitas XYZ. Setelah membahas jumlah peminat tiap tahunnya, penting untuk menganalisis pertumbuhan peminat di setiap program studi. Pemahaman terhadap pertumbuhan ini akan memberikan wawasan yang lebih mendalam mengenai dinamika minat calon mahasiswa dan membantu Universitas XYZ dalam mengidentifikasi program studi yang mengalami peningkatan atau penurunan signifikan. Berikut ini adalah analisis pertumbuhan peminat program studi dari tahun 2020 hingga 2023. Gambar 4. 18 Data Persentase Pertumbuhan 31 Tabel ini menunjukkan persentase peningkatan jumlah peminat mahasiswa baru pada berbagai program studi di Universitas XYZ. Pada tahun 2020, tidak ada data yang tersedia untuk semua program studi. Namun, pada tahun 2021, semua program studi mengalami peningkatan signifikan, dengan Desain Produk mencatat peningkatan tertinggi sebesar 227.27%, diikuti oleh Informatika (207.94%) dan Desain Komunikasi Visual (141.79%). Selanjutnya, pada tahun 2022, terjadi penurunan persentase peningkatan di beberapa program studi dibandingkan tahun sebelumnya. Meskipun demikian, Desain Produk tetap menunjukkan peningkatan tinggi sebesar 86.11%, sementara Teknik Sipil mencatat peningkatan terendah sebesar 22.47%. Pada tahun 2023, beberapa program studi mengalami peningkatan kembali, dengan Desain Produk mencatat peningkatan sebesar 61.19% dan Teknik Sipil meningkat menjadi 66.06%. 32

BAB V PENUTUP 5.1 Kesimpulan Atas dasar penelitian ini, metode

K-Nearest Neighbor (KNN) digunakan. untuk memprediksi peminatan program studi di Universitas XYZ dengan menggunakan data penerimaan mahasiswa baru dari tahun 2020 hingga 2023. Beberapa kesimpulan penting dapat ditarik dari temuan studi dan analisis data: - Dari analisis jumlah peminat tiap tahunnya, ditemukan bahwa program studi Ilmu Komunikasi, Manajemen, dan Psikologi consistently menempati posisi teratas dalam hal jumlah peminat. Ini menunjukkan preferensi yang kuat dari calon mahasiswa terhadap program studi tersebut. - Analisis pertumbuhan peminat menunjukkan fluktuasi signifikan di beberapa program studi. Desain Produk dan Informatika mencatat peningkatan yang tinggi, sementara Teknik Sipil mengalami pertumbuhan yang lebih moderat. - Hasil penerapan metode K-Nearest Neighbor (KNN) menunjukkan bahwa model ini memiliki tingkat akurasi sebesar 88%. Ini berarti model prediksi yang dikembangkan cukup kuat dan dapat diandalkan untuk memprediksi peminatan program studi di masa depan. - Visualisasi data membantu memberikan gambaran yang jelas tentang jumlah dan pertumbuhan peminat program studi dari tahun ke tahun, yang berguna untuk perencanaan dan pengambilan keputusan di Universitas XYZ. Singkatnya, penelitian ini dengan jelas menggambarkan bagaimana teknik K-Nearest Neighbor dapat digunakan untuk meramalkan secara akurat peminatan program studi, dengan akurasi yang mendekati hasil yang kuat. Data dan analisis yang diperoleh dapat dijadikan dasar untuk strategi peningkatan dan pengembangan program studi di Universitas XYZ.

5.2 Saran Dari hasil penelitian ini, beberapa saran dapat diajukan untuk memperbaiki penelitian di masa depan dan memanfaatkan hasil penelitian oleh Universitas XYZ: - Disarankan untuk memperluas cakupan data dengan memasukkan lebih banyak tahun akademik dan variabel tambahan, seperti latar belakang pendidikan calon mahasiswa, preferensi karir, dan faktor ekonomi. Hal ini dapat meningkatkan akurasi prediksi dan memberikan wawasan yang lebih komprehensif. - Selain K-Nearest Neighbor, metode lain seperti Decision Tree, Random Forest, atau Neural Networks dapat diuji dan dibandingkan untuk melihat apakah ada peningkatan akurasi

dan kinerja prediksi. - Disarankan untuk melakukan monitoring dan evaluasi berkala terhadap model prediksi yang digunakan. Data terbaru harus terus dimasukkan untuk menjaga relevansi dan akurasi prediksi. - Disarankan untuk mengembangkan aplikasi berbasis web atau mobile yang dapat memanfaatkan model prediksi KNN untuk memberikan rekomendasi program studi kepada calon mahasiswa. Aplikasi ini dapat membantu calon mahasiswa membuat keputusan yang lebih informasi berdasarkan data historis. 33 - Berdasarkan hasil analisis, Universitas XYZ dapat memperkuat strategi promosi untuk program studi yang memiliki tren peningkatan peminat yang stabil, serta melakukan evaluasi dan perbaikan pada program studi yang mengalami penurunan minat. - Bagian yang bertanggung jawab atas pengumpulan data disarankan untuk meningkatkan ketajaman dan cakupan data yang diperoleh, misalnya melalui program kerja yang lebih terstruktur dan terencana. Data yang lebih lengkap dan akurat dapat memberikan hasil prediksi yang lebih baik. - Penelitian lanjutan sebaiknya dilakukan untuk mengkaji faktor-faktor lain yang mempengaruhi peminatan program studi, termasuk pengaruh media sosial, reputasi program studi, dan testimoni dari alumni. Diharapkan saran-saran ini dapat diterapkan oleh Universitas XYZ agar lebih efektif dalam merencanakan dan mengembangkan program studi, serta meningkatkan daya tarik



REPORT #21945037

Results

Sources that matched your submitted document.

● IDENTICAL ● CHANGED TEXT

INTERNET SOURCE		
1.	0.42% dataminigkrisnopratama.blogspot.com http://dataminigkrisnopratama.blogspot.com/2017/10/arsitektur-model-data-m..	●
INTERNET SOURCE		
2.	0.39% ejournal.itn.ac.id https://ejournal.itn.ac.id/index.php/jati/article/download/8193/4847/	●
INTERNET SOURCE		
3.	0.32% nabiladnjsmn.medium.com https://nabiladnjsmn.medium.com/data-warehouse-memaksimalkan-peluang-k..	●
INTERNET SOURCE		
4.	0.29% ejournal.uin-suska.ac.id https://ejournal.uin-suska.ac.id/index.php/RMSI/article/download/24948/9630	●
INTERNET SOURCE		
5.	0.27% media.neliti.com https://media.neliti.com/media/publications/141759-ID-educational-data-minin...	●
INTERNET SOURCE		
6.	0.26% media.neliti.com https://media.neliti.com/media/publications/224659-penggunaan-data-mining-...	●
INTERNET SOURCE		
7.	0.23% media.neliti.com https://media.neliti.com/media/publications/261197-perbandingan-algoritma-k...	●
INTERNET SOURCE		
8.	0.2% www.slideshare.net https://www.slideshare.net/slideshow/chapter-2-fitur-dan-komponen-dataware...	●
INTERNET SOURCE		
9.	0.2% www.guru99.com https://www.guru99.com/id/data-mining-tutorial.html	●



REPORT #21945037

INTERNET SOURCE		
10. 0.2%	repository.its.ac.id https://repository.its.ac.id/43241/1/2713100130_Undergraduate_Theses.pdf	● ●
INTERNET SOURCE		
11. 0.19%	id.wikipedia.org https://id.wikipedia.org/wiki/Python_(bahasa_pemrograman)	●
INTERNET SOURCE		
12. 0.19%	repo.darmajaya.ac.id http://repo.darmajaya.ac.id/15676/9/BAB%20IV.pdf	●
INTERNET SOURCE		
13. 0.18%	e-journal.uajy.ac.id http://e-journal.uajy.ac.id/1899/2/1KOM02805.pdf	●
INTERNET SOURCE		
14. 0.18%	ejournal.itn.ac.id https://ejournal.itn.ac.id/index.php/jati/article/view/2656	●
INTERNET SOURCE		
15. 0.17%	repository.um-surabaya.ac.id https://repository.um-surabaya.ac.id/4178/4/BAB_III-1.pdf	●
INTERNET SOURCE		
16. 0.17%	elibrary.unikom.ac.id https://elibrary.unikom.ac.id/id/eprint/7475/8/UNIKOM_Brenda%20Casa%20Iql...	●
INTERNET SOURCE		
17. 0.16%	dqlab.id https://dqlab.id/simak-contoh-visualisasi-dalam-jenis-data-statistik	●
INTERNET SOURCE		
18. 0.16%	pasla.jambiprov.go.id https://pasla.jambiprov.go.id/business-intelligence-pengertian-manfaat-dan-tah..	●
INTERNET SOURCE		
19. 0.16%	eskripsi.usm.ac.id https://eskripsi.usm.ac.id/files/skripsi/G21A/2015/G.241.15.0001/G.241.15.0001-...	●
INTERNET SOURCE		
20. 0.15%	herza.id https://herza.id/blog/operating-system-definisi-fungsi-dan-jenisnya/	●



REPORT #21945037

INTERNET SOURCE		
21.	0.14% repository.yudharta.ac.id https://repository.yudharta.ac.id/351/1/BAB%20I.pdf	●
INTERNET SOURCE		
22.	0.14% badr.co.id https://badr.co.id/data-warehouse/	●
INTERNET SOURCE		
23.	0.14% ojs.unikom.ac.id https://ojs.unikom.ac.id/index.php/jati/article/download/10525/3908/	●
INTERNET SOURCE		
24.	0.13% esairina.medium.com https://esairina.medium.com/algoritma-k-nearest-neighbor-knn-penjelasan-dan..	●
INTERNET SOURCE		
25.	0.13% myedusolve.com https://myedusolve.com/id/blog/cara-cepat-untuk-menghapus-data-duplikat-e...	●
INTERNET SOURCE		
26.	0.13% seminar.iaii.or.id https://seminar.iaii.or.id/index.php/SISFOTEK/article/download/273/241	●
INTERNET SOURCE		
27.	0.13% jurnal.darmajaya.ac.id https://jurnal.darmajaya.ac.id/index.php/PSND/article/download/3848/1655	●
INTERNET SOURCE		
28.	0.12% medium.com https://medium.com/@myskill.id/python-classification-9310d427a0c1	●
INTERNET SOURCE		
29.	0.12% journal.unpas.ac.id https://journal.unpas.ac.id/index.php/infomatek/article/download/11209/6315/...	●
INTERNET SOURCE		
30.	0.11% www.trivusi.web.id https://www.trivusi.web.id/2022/06/algoritma-knn.html	●
INTERNET SOURCE		
31.	0.11% pustaka.ut.ac.id https://pustaka.ut.ac.id/lib/wp-content/uploads/pdfmk/MSIM4315-M1.pdf	●



REPORT #21945037

INTERNET SOURCE		
32.	0.1% sekolahstata.com https://sekolahstata.com/5-teknik-untuk-meminimalkan-kesalahan-dalam-pene..	●
INTERNET SOURCE		
33.	0.09% mysertifikasi.com https://mysertifikasi.com/langkah-langkah-mengatasi-missing-values-pada-dat...	●
INTERNET SOURCE		
34.	0.09% www.diklatkerja.com https://www.diklatkerja.com/blog/memahami-algoritma-k-nearest-neighbor-k-...	●
INTERNET SOURCE		
35.	0.09% www.journal-computing.org https://www.journal-computing.org/index.php/journal-sea/article/download/12...	●
INTERNET SOURCE		
36.	0.06% repository.unas.ac.id http://repository.unas.ac.id/6713/3/BAB%202.pdf	●
INTERNET SOURCE		
37.	0.05% jdih.kemenkeu.go.id https://jdih.kemenkeu.go.id/fulltext/2012/12TAHUN2012UU.htm	●
INTERNET SOURCE		
38.	0.05% digi-lib.stekom.ac.id https://digi-lib.stekom.ac.id/assets/dokumen/ebook/feb_BMqCPd7pUQAohZqd...	●

● QUOTES

INTERNET SOURCE		
1.	0.49% e-journal.uajy.ac.id http://e-journal.uajy.ac.id/1899/2/1KOM02805.pdf	
INTERNET SOURCE		
2.	0.34% jdih.kemenkeu.go.id https://jdih.kemenkeu.go.id/fulltext/2012/12TAHUN2012UU.htm	
INTERNET SOURCE		
3.	0.16% ejournal.uin-suska.ac.id https://ejournal.uin-suska.ac.id/index.php/RMSI/article/download/24948/9630	

REPORT #21945037

INTERNET SOURCE

4. **0.14%** ejournal.itn.ac.id

<https://ejournal.itn.ac.id/index.php/jati/article/view/2656>