Isian Substansi Proposal

SKEMA PENELITIAN DASAR

Petunjuk:Pengusul hanya diperkenankan mengisi di tempat yang telah disediakan sesuai dengan petunjuk pengisian dan tidak diperkenankan melakukan modifikasi template atau penghapusan di setiap bagian.

JUDUL

Tuliskan Judul Usulan

Penerapan Data Mining dengan Teknik Klasifikasi untuk Mengidentifikasi Faktor-faktor yang Mempengaruhi Siswa SMK Ashkabul Kahfi Keluar atau Dimutasi.

RINGKASAN

Ringkasan penelitian tidak lebih dari 300 kata yang berisi urgensi, tujuan, dan luaran yang ditargetkan.

Urgensi. SMK Askhabul Kahfi adalah salah satu SMK swasta yang ada di Indonesia. SMK Askhabul Kahfi memiliki masalah dengan siswa yang dikeluarkan atau dimutasi. Data dari SMK Askhabul Kahfi menunjukkan bahwa ada sekitar 20.86% siswa yang dikeluarkan atau dimutasi setiap tahunnya. Keluar atau dimutasi siswa SMK Askhabul Kahfi dapat berdampak negatif terhadap kualitas lulusan SMK Askhabul Kahfi. Siswa yang dikeluarkan atau dimutasi mungkin tidak memiliki keterampilan dan pengetahuan yang memadai untuk bekerja. Hal ini dapat menurunkan daya saing lulusan SMK Askhabul Kahfi di dunia kerja. **Tujuan**. Dari urgensi yang telah penulis uraikan, dapat diasumsikan bahwa SMK Askhabul Kahfi membutuhkan analisis data untuk meningkatkan kualitas SMK Askhabul Kahfi. Oleh karena itu, dilakukan penelitian untuk menemukan faktor-faktor yang mempengaruhi siswa untuk dikeluarkan atau dimutasi. Penelitian ini diharapkan dapat memberikan informasi yang bermanfaat untuk mencegah siswa dikeluarkan atau dimutasi. Metode penelitian ini menggunakan metode data mining dengan teknik klasifikasi. Data yang digunakan adalah data alumni SMK Askhabul Kahfi. Data alumni SMK Askhabul Kahfi terdiri dari data demografi dan data ekonomi keluarga. Metode data mining dengan teknik klasifikasi akan diguankan untuk menemukan faktor-faktor yang mempengaruhi siswa untuk dikeluarkan atau dimutasi. Faktor-faktor tersebut akan diklasifikasikan ke dalam dua kategori, yaitu faktor internal dan faktor eksternal. **Luaran**. Luaran yang akan dihasilkan dalam proses klasifikasi ini adalah sebuah aplikasi berbasis web. Dimana terdapat tampilan UI untuk melihat data yang sudah di proses klasifikasi dengan metode yang sudah ditentukan.

KATA KUNCI

Kata kunci maksimal 5 kata

SMK Ashkabul Kahfi; Data Mining; Klasifikasi; Faktor yang mempengaruhi siswa dikeluarkan

PENDAHULUAN

Penelitian Dasar merupakan riset yang memuat temuan baru atau pengembangan ilmu pengetahuan dari kegiatan riset yang terdiri dari tahapan penentuan asumsi dan dasar hukum yang akan digunakan, formulasi konsep dan/ atau aplikasi formulasi dan pembuktian konsep fungsi dan/ atau karakteristik penting secara analitis dan eksperimental.

Pendahuluan penelitian tidak lebih dari 1000 kata yang terdiri dari:

A. Latar belakang dan rumusan permasalahan yang akan diteliti
B. Pendekatan pemecahan masalah
C. *State of the art* dan kebaruan
D. Peta jalan (*road map*) penelitian 5 tahun kedepan (jika dalam bentuk konsorsium harus dilengkapi dengan roadmap penelitian konsorsium)
E. Sitasi disusun dan ditulis berdasarkan sistem nomor sesuai dengan urutan pengutipan, mengikuti format Vancouver

Latar Belakang. SMK Askhabul Kahfi Semarang adalah sekolah lanjutan tinggi yang berada dibawah yayasan Nurul Ittifaq, sekolah ini mencetak siswa menjadi tenaga terampil dan siap untuk bekerja atau berwirausaha secara mandiri. SMK Askhabul Kahfi berlokasi di Jalan Cangkiran-gunungpati km.3 Polaman, Kec Mijen, Semarang. Dalam Pembelajarannya, SMK Askhabul Kahfi memadukan kurikulum Dinas Pendidikan dan kurikulum pesantren. Sehingga disamping belajar pelajaran dan teknik sebagaimana sekolah-sekolah pada umumnya, Santri SMK Askhabul Kahfi juga mempelajari ilmu-ilmu agama ala pesantren di kelas. Ilmu seperti fiih, tasawuf, nahwu, shorof, dll masuk sebagai pelajaran sekolah dengan kitab kuning sebagai bahan ajarnya. Meskipun dalam satu lembaga, Santri Putra dan putri Askhabul kahfi terpisah dalam dua lokal. Kampus 2 bagi santri putra dan kampus 3 bagi santri putri. Baik dalam kelas, gedung sekolah, maupun asrama dan tempat mengaji. SMK Askhabul Kahfi merupakan sebuah sekolah menengah kejuruan swasta di Indonesia .

Salah satu permasalahan yang dihadapi sekolah ini adalah masih terdapat peserta didik yang mengalamani pemutusan sekolah dengan alasan mutasi, dikeluarkan, mengundurkan diri, atau pemutusan sekolah dengan alasan lain. Berdasarkan wawancara penulis dengan staff Tata Usaha, mayoritas di keluarkannya siswa adalah karena kenakalan siswa yang tidak bisa ditoleransi. Namun berdasarkan data dari SMK Askhabul Kahfi menunjukkan bahwa setiap tahun sekitar 20.86%siswa mengalami pengeluaran atau pemutusan. Dampak dari masalah ini adalah potensi menurunnya kualitas lulusan SMK Askhabul Kahfi, karena siswa yang mengalami pengeluaran atau pemutusan mungkin tidak memiliki keterampilan dan pengetahuan yang memadai untuk bekerja. Skill apapun dapat dipelajari namun membutuhkan dedikasi yang kuat untuk mempelajari ilmu tersebut seperti perlunya mental positif, semangat motivasi, waktu dan terkadang uang. Sedangkan Knowledge (pengetahuan) adalah kemampuan seseorang untuk mengenali suatu keadaan berdasarkan persepsi pikirannya. Knowledge seseorang ditentukan oleh apa yang dipelajari dari bahan bacaan, lingkungan pergaulan, pekerjaan dan lain sebagainya. (1).

Penelitian ini bertujuan untuk mengidentifikasi faktor-faktor yang memengaruhi siswa keluar atau dikeluarkan. Dengan pemahaman yang lebih baik tentang faktor-faktor ini, langkah-langkah preventif dapat diambil untuk mengurangi tingkat keluarnya siswa di SMK Askhabul Kahfi dengan alasan selain lulus atau *dropout* sebelum lulus.

Rumusan Masalah. Penelitian ini menggunakan metode data mining dengan teknik klasifikasi. Data yang digunakan adalah data alumni SMK Askhabul Kahfi yang mencakup data demografi dan data ekonomi keluarga. Metode data mining akan membantu mengidentifikasi pola dan hubungan antara variabel-variabel ini dengan pengeluaran atau pemutusan siswa. Metode penelitian propoal ini menggunakan teknik klasifikasi. Berdasarkan hasil penelitian yang dilakukan oleh para penulis jurnal "*A Study of Some Data Mining Classfification Tecniques",* teknik klasifikasi yang paling akurat adalah Support Vector Machines. Namun, teknik klasifikasi Naive Bayes memiliki kinerja yang lebih baik dalam hal kecepatan dan kemudahan penggunaan.

(1). Namun jika dibandingkan dengan jurnal berjudul "*Analysis of Various Decision Tree Algorithms for Classification in Data Mining*" (2). Algoritma C4.5 yang paling baik, dalam jurnal tersebut dianalisis 4 algoritma yaitu ID3, C4.5, CART dan CAHD. Dari ke-empat algoritma dianalisis kinerjanya, hasil akhirnya didapatkan bahwa hasil penelitian yang dilakukan oleh para penulis jurnal tersebut, algoritma pohon keputusan C4.5 memiliki kinerja yang paling akurat.

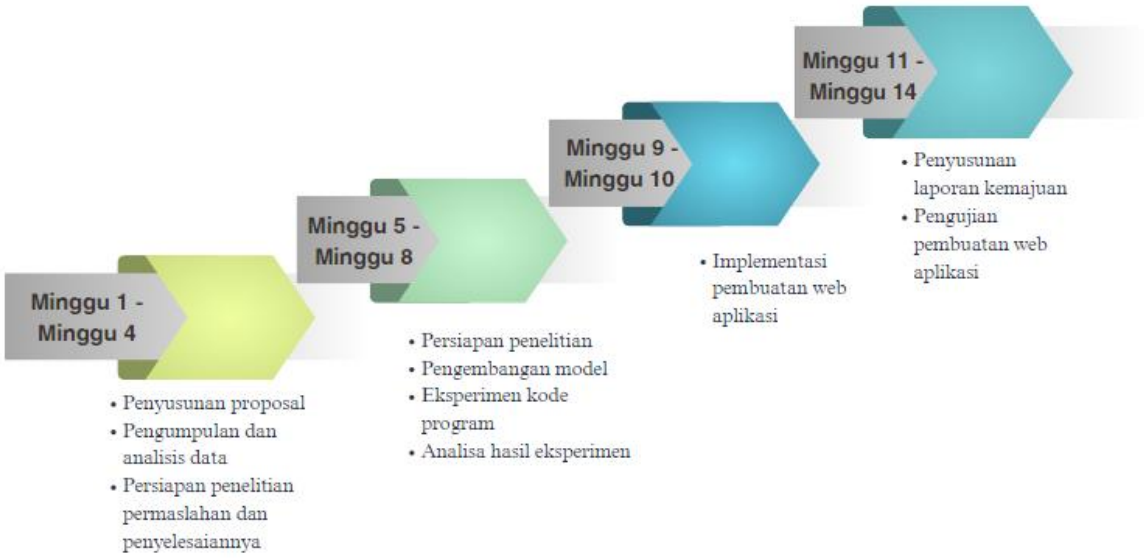
Berdasarkan analisis yang sudah dilakukan oleh beberapa peneliti pada dua jurnal yang sudah penulis sebutkan sebelumnya. Maka dari itu, penulis melakukan percobaan untuk algoritma pohon keputusan C4.5, Selain itu penulis juga melakukan percobaan dengan 5 kode program yang berbeda dengan 3 teknik yang digunakan yaitu, Decision Tree, Naive Bayes, dan K-Nearest Neighbord (K-NN).

State of the Art dan Kebaharuan. Penelitian ini merupakan pengembangan dari penelitian-penelitian sebelumnya dalam bidang data mining untuk memprediksi pengeluaran atau pemutusan siswa. Namun, penelitian ini memiliki kebaruan dalam pendekatan penggunaan teknik data mining dengan teknik klasifikasi khusus untuk SMK Askhabul Kahfi. Data yang digunakan dalam penelitian ini juga merupakan data alumni yang spesifik untuk sekolah ini.

Dari penelitian tersebut, penulis telah melakukan percobaan dengan mengklasfiikasikan data alumni SMK Ashkabul Kahfi dengan lima kode program/algoritma yang berbeda, yaitu kode program dengan teknik decision tree (2 kode program/algoritma berbeda), naïve bayes (2 kode program/algoritma yang berbeda) dan dengan teknik K-NN (satu algoritma).

Peta Jalan (*Road Map*) 14 Minggu. Penelitian ini akan dilakukan selama 14 minggu. Pada minggu pertama, kedua, ketida dan keempat fokus akan diberikan pada tahap pemahaman data, persiapan data, dan pemodelan. Pada minggu lima hingga minggu delapan penelitian akan berfokus pada tahap evaluasi dan implementasi. Minggu ke sembilan dan sepuluh selama dua minggu akan difokuskan untuk membangun luaran yang berupa web aplikasi menggunakan django atau steamlit, hingga setelah jika web aplikasi sudah selesai maka akan dilanjut proses pengujian web aplikasi. Pada minggu terkahir, minggu ke sebelas hingga empat bleas akan dilakukan review tulis laporan penelitian yang mencakup seluruh proses dari identifikasi inovasi hingga hasil analisis dan juga pengujian untuk hasil web aplikasi yang sudah dibangun (luaran).

Dari peta jalan ini, kita dapat mendapat gambaran bagaimana dapat menghasilkan sebuah produk dengan data mining dengan teknik klasifikasi. Secara garis besar *road map* bisa dilihat pada gambar 1.1 road map penelitian dibawah.



Gambar 1. 1 Road Map Penelitian

Tabel 1. 1 Rincian Road Map

Minggu Ke-	Kegiatan	Keterangan
1 - 4	- Penyusunan Proposal - Pengumpulan dan analisis data - Persiapan penelitian permasalahan dan penyelesaiannya	Pada tahap ini, rencananya adalah dengan membaca proposal awal dengan seksama. Memastikan bahwa semua detailnya dan identifikasi inovasi yang bisa dikembangka
		Analisis kebutuhan: Identifikasi apakah inovasi diperlukan untuk menjawab pertanyaan penelitian. Kemudian pengumpulan data.
5 - 8	- Persiapan penelitian - Pengembangan model - Eksperimen kode program/algoritma	Pada minggu ke 5 – 8, setelah mendapatkan data mentahan yang akan diolah akan dilakukan persiapan data. Dari data mentahan akan di preprocessing agar data mentahan siap untuk diuji/diproses. Kemudian pada minggu ini dilakukan juga sebuah eksperimen untuk beberapa algoritma, yang dimana pengujian dilakukan menggunakan 3 metode (decision tree, naïve bayes, K-NN).
9 – 10	- Implementasi pembuatan web aplikasi (luaran)	Pada minggu ini akan dimulainya pengerjaan implementasi pembuatan web aplikasi (hasil luaran) produk klasifikasi untuk mengidentifikasi faktor-faktor dikeluarkannya atau mutasi murid. Implementasi web aplikasi akan menggunakan framework django atau steamlit.
11 – 14	- Penyusunan laporan kemajuan - Pengujian pembuatan web aplikasi	Di tahap minggu terakhir, minggu ke-11 hingga minggu ke-14 dimana minggu ini targetnya adalah aplikasi web sudah selesai dibangun. Maka, akan dilakukan sebuah pengujian kelayakan dan akurasi web aplikasi.

METODA

Metode atau cara untuk mencapai tujuan yang telah ditetapkan ditulis tidak melebihi 1000 kata. Bagian ini dapat dilengkapi dengan diagram alir penelitian yang menggambarkan apa yang sudah dilaksanakan dan yang akan dikerjakan selama waktu yang diusulkan. Format diagram alir dapat berupa file JPG/PNG. Metode penelitian harus dibuat secara utuh dengan penahapan yang jelas, mulai dari awal bagaimana proses dan luarannya, dan indikator capaian yang ditargetkan yang tercermin dalam Rencana Anggaran Biaya (RAB).

Data Mining, Klasfikasi, dan Decision Tree

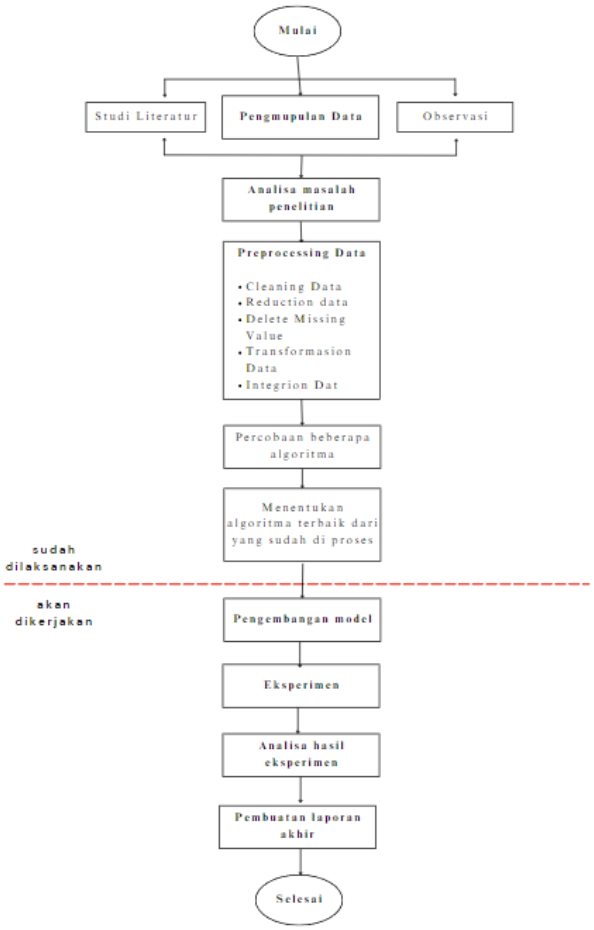
Data mining adalah serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Informasi yang dihasilkan dapat diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat dalam basis data (1). Data mining juga banyak meliputi banyak bidang ilmu contohnya artificial intelligent, machine learning, statistic dan basis data (2). Data mining juga digunakan untuk menentukan pola serta informasi dalam suatu data (2).

Salah satu teknik untuk mengolah data di data mining adalah klasifikasi. Klasifikasi adalah fungsi penambangan data yang menetapkan item dalam koleksi ke kategori atau kelas target. Tujuan klasifikasi adalah untuk memprediksi secara akurat kelas target untuk setiap kasus dalam data. Misalnya, model klasifikasi dapat digunakan untuk mengidentifikasi pemohon pinajaman berdasarkan risiko kredit rendah, menengah, atau tinggi. Tugas klasifikasi dimulai dengan kumpulan data yang penetapan kelasnya diketahui (4). Dari data yang sudah penulis peroleh, terdapat target yaitu ‘Status Keluar’. Pada atribut ‘Status Keluar’ berisi beberapa kategori yaitu, “lulus”, “dikeluarkan”, “mutasi”, “mengundurkan diri”, dan “lainnya”. Dari atribut dan isi dari atribut tersebut dapat diambil data sebagai target.

Langkah selanjutnya adalah menemukan metode untuk pengklasifkasiannya sendiri. Konsep yang digunakan dalam karya ini untuk teknik klasifikasi kelulusan siswa adalah

decision tree (pohon keputusan). Pohon keputusan adalah struktur seperti pohon dimana simpul internal berisi atribut pemisahan dan pemisahan. Itu mewakili tes pada atributnya. Phon keputusan biasanya dibangun dari kumpulan data pelatihan sedangkan kumpulan data pengujian dibuat diguankan untuk menguji atau memvalidasi keakuratan pohon keputusan. Phon keputusan adalah struktur pohon seperti diagram alur yang memiliki ciri-ciri berikut: (1) setiap simpul internal juga disebut sebagai simpul non-daun menunjukkan pengujian pada satu atribut; (2) setiap cabang mewaiki hasil ujian; (3) setiap simpul daun atau simpul terminal memiliki kelas label; (4) simpul paling atas dari pohon adalah simpul akar. Keputusan biasanya terdiri dari simpul-simpul yang membentuk a pohon berakar, yang berarti pohon berarah dengan simpul yang disebut akar yang tidak mempunyai tepi masuk.

Selanjutnya untuk memulai proses data mining menggunakan klasifikasi dengan decision tree dibutuhkan adanya pemrosesan awal data. Dibutuhkan juga tahapan penelitian untuk menjadi acuan dalam penelitian. Dalam praktiknya, tahap penelitian proposal ini adalah sebagai berikut.



Gambar 1. 2 Tahapan Penelitian

1. **Pengumpulan Data**

Teknik pengumpulan data merupakan teknik yang digunakan dalam mengumpulkan suatu informasi terkait yang dibuthkan dalam melakukan penelitian. Adapun teknik pengumpulan data yang digunakan dalam penelitian ini yaitu:

a. Observasi

Metode ini dilakukan dengan secara langsung pada objek yang akan diteliti di SMK Askhabul Kahfi secara langsung untuk mendapatkan data siswa lulus di tahun sebelumnya. Berdasarkan data yang sudah didapatkan dengan menjelaskan tujuan dari data tersebut, dilakukanlah sebuah observasi terhadap data.

b. Studi Literaur

Studi Literatur dilakukan untuk meningkatkan kemampuan peneliti dalam memahami dan mengetahui tren studi yang berkaitan dengan masa belajar atau sekolah SMK di Indonesia. Atau untuk mengetahu bagaimana cara kerja sebuah kodingan dengan teknik algoritma python bisa berpengaruh di dalam kasus ini.

Berikut data mentahan daftar siswa yang keluar dari SMK Askhabul Kahfi dengan status lulus, dikeluarkan, mutasi, atau lainnya.

Data Ayah				Data Ibu						Data Wali						
njang Pendidik	Pekerjaan	Penghasilan	ebutuhan Khus	Nama	Tahun Lahir	njang Pendidik	Pekerjaan	Penghasilan	ebutuhan Khus	Nama	Tahun Lahir	njang Pendidik	Pekerjaan	Penghasilan	Keluar Karena	Tanggal keluar
SMP / sederajat	Wiraswasta	Rp. 5,000,000 - R	Tidak ada	NI'MAH AWALIY	1976	D3	Karyawan Swast	Rp. 2,000,000 - R	Tidak ada		0				Lulus	2021-07-01
SMA / sederajat	Karyawan Swast	Rp. 500,000 - Rp	Tidak ada	RAHAYU MULYA	1972	SMP / sederajat	Karyawan Swast	Rp. 2,000,000 - R	Tidak ada		0	Tidak sekolah			Lulus	2023-07-01
SD / sederajat	Karyawan Swast	Rp. 1,000,000 - R	Tidak ada	Sri Indarti	1976	SMP / sederajat	Tidak bekerja	Tidak Berpengha	Tidak ada		0				Mutasi	2023-01-19
SD / sederajat	Buruh	Rp. 1,000,000 - R	Tidak ada	SUNAH	1958	SD / sederajat	Buruh	Rp. 1,000,000 - R	Tidak ada	-	1900				Lulus	2016-05-06
SD / sederajat	Petani		Tidak ada	SUMTIAH	1965	SD / sederajat	Petani		Tidak ada		0	Lainnya	Lainnya		Lulus	2015-05-18
Putus SD	Pensiunan	Rp. 1,000,000 - R	Tidak ada	NUR ASIH	1969	Putus SD	Tidak bekerja		Tidak ada	-	1900				Lulus	2017-05-03
SMA / sederajat	Karyawan Swast	Rp. 1,000,000 - R	Tidak ada	Pujiati		SMP / sederajat	Petani	Rp. 1,000,000 - R	Tidak ada						Lulus	2021-07-01
SMP / sederajat	Buruh		Tidak ada	PARWIYATI	1969	SD / sederajat	Tidak bekerja		Tidak ada		0	Lainnya	Lainnya		Lulus	2015-05-18
SD / sederajat	Petani		Tidak ada	SOPIYAH	1972	SD / sederajat	Petani		Tidak ada		1900				Lulus	2015-05-18
Putus SD	Buruh	Rp. 1,000,000 - R	Tidak ada	RUKINI	1968	Putus SD	Tidak bekerja		Tidak ada	-	1900				Lulus	2016-05-06
SD / sederajat	Petani		Tidak ada	YATIMAH	1975	SD / sederajat	Petani		Tidak ada		1900				Lulus	2015-05-18
SMP / sederajat	Wiraswasta	Rp. 1,000,000 - R	Tidak ada	NUR HIDAYAH	1972	SMP / sederajat	Wiraswasta	Rp. 500,000 - Rp	Tidak ada	-	1900				Lulus	2017-05-02
SMP / sederajat	Wiraswasta	Rp. 1,000,000 - R	Tidak ada	JAMI	1979	SD / sederajat	Wiraswasta	Rp. 500,000 - Rp	Tidak ada	-	1900				Lulus	2018-07-01
SMA / sederajat	Buruh	Rp. 1,000,000 - R	Tidak ada	SRI REJEKI	0	SMP / sederajat	Karyawan Swast	Rp. 1,000,000 - R	Tidak ada		0	(tidak diisi)			Mutasi	2019-08-26
SD / sederajat	Petani		Tidak ada	SOELASMI	1958	SD / sederajat	Petani		Tidak ada		1900				Lulus	2015-05-18
SMP / sederajat	Petani	Rp. 1,000,000 - R	Tidak ada	SUHARTI	0	SMP / sederajat	Petani	Rp. 1,000,000 - R	Tidak ada						Dikeluarkan	2021-09-17
SMP / sederajat	Petani	Rp. 1,000,000 - R	Tidak ada	SUHARTI	0	SMP / sederajat	Petani	Rp. 1,000,000 - R	Tidak ada						Dikeluarkan	2021-09-17
SMA / sederajat	Wiraswasta	Rp. 500,000 - Rp	Tidak ada	EPI SOPARIYATI	1978	SMA / sederajat	Tidak bekerja		Tidak ada	-	1900				Lulus	2016-05-06
			Tidak ada	KHUSNUL ULFAH					Tidak ada						Dikeluarkan	2019-10-29
SMA / sederajat	Petani	Rp. 1,000,000 - R	Tidak ada	KHUSNUL ULFAH	1984	D3	Petani	Rp. 1,000,000 - R	Tidak ada		0	Tidak sekolah			Lulus	2022-07-01
			Tidak ada	KHUSNUL ULFAH					Tidak ada						Dikeluarkan	2019-10-29
SMP / sederajat	Wiraswasta	Rp. 1,000,000 - R	Tidak ada	Marfuah	1983	SMP / sederajat	Wiraswasta	Rp. 1,000,000 - R	Tidak ada		0				Mutasi	2020-11-16
SMP / sederajat	Karyawan Swast	Rp. 1,000,000 - R	Tidak ada	TUTIK	1977	SMA / sederajat	Karyawan Swast	Rp. 1,000,000 - R	Tidak ada		0	Tidak sekolah			Lulus	2021-07-01
SMP / sederajat	Buruh	Rp. 1,000,000 - R	Tidak ada	Salbiyah		SD / sederajat	Petani	Rp. 1,000,000 - R	Tidak ada						Lulus	2021-07-01
SD / sederajat	Petani		Tidak ada	NIKMAH IZATI	1969	SD / sederajat	Petani		Tidak ada		1900				Lulus	2015-05-18
SMP / sederajat	Petani	Rp. 1,000,000 - R	Tidak ada	CHOFIFAH	0	SMP / sederajat	Wiraswasta	Rp. 1,000,000 - R	Tidak ada						Lulus	2023-07-01
	Sudah Meninggal		Tidak ada	ELA SITI NGATIY	1978	SD / sederajat	Petani		Tidak ada		1900				Lulus	2015-05-18
SMA / sederajat	Petani	Rp. 500,000 - Rp	Tidak ada	RUSTIYAH	1971	SD / sederajat	Karyawan Swast	Rp. 1,000,000 - R	Tidak ada	-	1900				Lulus	2018-07-01
SD / sederajat	Petani	Rp. 1,000,000 - R	Tidak ada	SULNIYATI	1985	SD / sederajat	Petani	Rp. 1,000,000 - R	Tidak ada		0	(tidak diisi)	Petani		Lulus	2023-07-01
SMA / sederajat	Karyawan Swast	Rp. 1,000,000 - R	Tidak ada	YULIYANTI	1980	SMA / sederajat	Tidak bekerja	Kurang dari Rp.	Tidak ada		0	Tidak sekolah			Lulus	2020-07-01

Gambar 1. 3 Data Mentahan Alumni SMK Askhabul Kahfi

2. Analisa Masalah Penelitian

Analisa masalah penelitian ini dilakukan untuk memahami lebih lanjut tentang masalah yang dihadapi oleh SMK Askhabul Kahfi, yaitu adanya siswa yang dikeluarkan atau dimutasi. Analisa masalah penelitian ini dilakukan dengan menggunakan metode data mining python dan algoritma yang dapat digunakan untuk klasifikasi, yaitu decision tree, naïve bayes, dan k-nn. Berdasarkan data yang sudah didapatkan, berisi beberapa data diri siswa secara demografis dan data diri orang tua mereka beserta status haji mereka.

Dari data yang sudah didapatkan, dapat direkap bahwa terdapat 1036 *record* data, sebanyak 43 atribut yang dapat dikelompokkan menjadi 3 bagian, yaitu data siswa itu sendiri, data ibu dan data ayah dari siswa.

Ketersediaan data yang memiliki banyak atribut dapat digunakan untuk mengklasifikasikan siswa yang keluar karena alasan dikeluarkan atau mutasi. Selanjutnya data mentahan akan dibersihkan di *preprocessing data*.

3. Preprocessing Data

Preprocessing data adalah proses untuk mempersiapkan data agar siap untuk dianalisis. Data yang kita miliki biasanya tidak dalam kondisi yang siap untuk dianalisis. Data yang kita miliki mungkin mengandung kesalahan, anomali, atau format yang tidak sesuai dengan kebutuhan analisis. Preprocessing data dapat membantu kita untuk mengatasi masalah-masalah tersebut agar hasil analisis yang kita dapatkan lebih akurat dan informatif.

a. Data Cleaning

Data cleaning adalah proses untuk menghilangkan kesalahan dan anomali dari data. Data cleaning dapat dilakukan dengan menggunakan berbagai teknik, seperti pengecekan keakuratan data, pengisian data yang hilang, dan penghapusan data yang tidak relevan.

Setelah mendapatkan data mentahan, dilakukan pembersihan data untuk yang tidak relevan. Berikut adalah potongan data yang sudah saya hilangkan kolom yang

tidak relevan dengan tujuan proses data mining. Pada proses ini data mentahan ada sebanyak 1036 record menjadi tetap karena tidak ada duplikasi record, namun untuk atribut di hapus menjadi tersisa hanya 13 atribut saja karena atribut diseleksi berdasarkan ‘apakah atribut ini dapat berpengaruh terhadap siswa keluar?’.

JK	Tanggal Lahir	Tempat Lahir	Asal Kecamatan	Jenis Tinggal	Alat Transportasi	Jenjang Pendidikan Ayh	Pekerjaan Ayh	Penghasilan Ayh	Jenjang Pendidikan Ibu	Pekerjaan Ibu	Penghasilan Ibu	Keluar Karena
L	2003-01-28	SEMARANG	Cilacap Tengah	Bersama orang tua	Sepeda	SMP / sederajat	Wiraswasta	Rp. 5,000,000 - Rp. 20,000,000	D3	Karyawan Swasta	Rp. 2,000,000 - Rp. 4,999,999	Lulus
L	2004-11-14	SEMARANG	Mijen	Bersama orang tua	Jalan kaki	SMA / sederajat	Karyawan Swasta	Rp. 500,000 - Rp. 999,999	SMP / sederajat	Karyawan Swasta	Rp. 2,000,000 - Rp. 4,999,999	Lulus
L	2006-05-26	Semarang	Ngaliyan	Bersama orang tua	Ojek	SD / sederajat	Karyawan Swasta	Rp. 1,000,000 - Rp. 1,999,999	SMP / sederajat	Tidak bekerja	Tidak Berpenghasilan	Mutasi
L	1998-07-08	TEGAL	Balapulang	Panti asuhan	Sepeda motor	SD / sederajat	Buruh	Rp. 1,000,000 - Rp. 1,999,999	SD / sederajat	Buruh	Rp. 1,000,000 - Rp. 1,999,999	Lulus
L	1995-12-18	SEMARANG	Semarang	Asrama	Jalan kaki	SD / sederajat	Petani	NaN	SD / sederajat	Petani	NaN	Lulus

Gambar 1. 4 Dataset Setelah di Bersihkan

Dalam kolom ‘Tanggal Lahir’ terdapat format date didalamnya, jika menggunakan format tersebut maka sulit diguankan untuk mendukung faktor siswa dropout. Maka dari itu, sisakan saja tahun lahir siswa agar katerogi lebih besar.

```
[ ]  # ubah format tanggal lahir dgn hanya tahun saja (agar lbh general)
     data['Tanggal Lahir'] = pd.to_datetime(data['Tanggal Lahir']).dt.year
```

```
[ ]  print(data)
```

```
     JK  Tanggal Lahir Tempat Lahir  Asal Kecamatan      Jenis Tinggal  \
0     L           2003     SEMARANG  Cilacap Tengah  Bersama orang tua
1     L           2004     SEMARANG           Mijen  Bersama orang tua
2     L           2006     Semarang        Ngaliyan  Bersama orang tua
3     L           1998        TEGAL      Balapulang       Panti asuhan
4     L           1995     SEMARANG        Semarang             Asrama
...  ..            ...          ...             ...                ...
1031  P           2002     SEMARANG   Ungaran Timur           Pesantren
1032  P           2006     Semarang        Podorejo  Bersama orang tua
1033  P           1997       KENDAL       Limbangan             Asrama
1034  P           2001     SEMARANG         Tuntang                NaN
1035  P           2003     Semarang        Ngaliyan             Asrama
```

Gambar 1. 5 Potongan kode preprocessing data mentahan

Proses selanjutnya yaitu memastikan bahwa tidak terdapat data yang kosong pada semua record. Berikut kode untuk cleaning data kosong.

```
[ ]  # menghitung jumlah missing value di setiap kolom
     data.isnull().sum()
```

```
JK                        0
Tanggal Lahir             0
Tempat Lahir              0
Asal Kecamatan            0
Jenis Tinggal            37
Alat Transportasi        29
Jenjang Pendidikan Ayh   42
Pekerjaan Ayh            42
Penghasilan Ayh         291
Jenjang Pendidikan Ibu   24
Pekerjaan Ibu            25
Penghasilan Ibu         320
Keluar Karena             0
dtype: int64
```

Gambar 1. 6 Potongan kode preprocessing data mentahan (2)

```
[ ]  # mengisi missing value dgn modulus
     data["Jenis Tinggal"].fillna(data["Jenis Tinggal"].mode().iloc[0], inplace=True)
     data["Alat Transportasi"].fillna(data["Alat Transportasi"].mode().iloc[0], inplace=True)
     data["Jenjang Pendidikan Ayh"].fillna(data["Jenjang Pendidikan Ayh"].mode().iloc[0], inplace=True)
     data["Penghasilan Ayh"].fillna(data["Penghasilan Ayh"].mode().iloc[0], inplace=True)
     data["Pekerjaan Ayh"].fillna(data["Penghasilan Ayh"].mode().iloc[0], inplace=True)
     data["Jenjang Pendidikan Ibu"].fillna(data["Jenjang Pendidikan Ibu"].mode().iloc[0], inplace=True)
     data["Pekerjaan Ibu"].fillna(data["Pekerjaan Ibu"].mode().iloc[0], inplace=True)
     data["Penghasilan Ibu"].fillna(data["Penghasilan Ibu"].mode().iloc[0], inplace=True)

     # setelah di preprocessing
     data.isnull().sum()
```

Gambar 1. 7 Potongan kode preprocessing data mentahan (3)

b. Data Transformation

Data transformation adalah proses untuk mengubah format data agar sesuai dengan kebutuhan analisis. Data transformation dapat dilakukan dengan menggunakan berbagai teknik, seperti konversi data ke format numerik, standarisasi data, dan normalisasi data.

Dari data yang sudah dibersihkan, sudah dipastikan tidak terdapat sel atau record yang kosong dapat dihilangkan dengan teknik factorize data mengguankan python.

```
[ ]  data["JK"] = pd.factorize(data.JK)[0]
     data["Tempat Lahir"] = pd.factorize(data["Tempat Lahir"])[0]
     data["Asal Kecamatan"] = pd.factorize(data["Asal Kecamatan"])[0]
     data["Jenis Tinggal"] = pd.factorize(data["Jenis Tinggal"])[0]
     data["Alat Transportasi"] = pd.factorize(data["Alat Transportasi"])[0]
     data["Jenjang Pendidikan Ayh"] = pd.factorize(data["Jenjang Pendidikan Ayh"])[0]
     data["Pekerjaan Ayh"] = pd.factorize(data["Pekerjaan Ayh"])[0]
     data["Penghasilan Ayh"] = pd.factorize(data["Penghasilan Ayh"])[0]
     data["Jenjang Pendidikan Ibu"] = pd.factorize(data["Jenjang Pendidikan Ibu"])[0]
     data["Pekerjaan Ibu"] = pd.factorize(data["Pekerjaan Ibu"])[0]
     data["Penghasilan Ibu"] = pd.factorize(data["Penghasilan Ibu"])[0]
     data["Keluar Karena"] = pd.factorize(data["Keluar Karena"])[0]

     # Save data
     data.to_csv("data_numerik_satu.csv")
```

Gambar 1. 8 Potongan kode preprocessing data mentahan (4)

c. Data Reduction

Data reduction adalah proses untuk mengurangi ukuran data agar lebih mudah untuk dianalisis. Data reduction dapat dilakukan dengan menggunakan berbagai teknik, seperti agregasi data, seleksi fitur, dan dimensionality reduction.

```
[ ]  # splitting the data
     from sklearn.model_selection import train_test_split

     # define input and target variable
     x = data[["JK", "Tempat Lahir", "Asal Kecamatan", "Jenis Tinggal", "Alat Transportasi", "Jenjang Pendidikan Ayh", "Pekerjaan Ayh", "Penghasilan Ayh", "Pekerjaan Ibu", "Penghasilan Ibu"]]
     y = data["Keluar Karena"]

     # splitting test and training data
     x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=42)
```

Gambar 1. 9 Potongan kode preprocessing data mentahan (5)

4. Percobaan Beberapa Algorima

Pilihan algoritma pembelajaran spesifik mana yang harus kita gunakan merupakan langkah penting. Pengklasifikasian ditentukan oleh label pengklasifikasi (pemetaan dari instance yang tidak berlabel ke kelas) (4).

Saya telah mencoba menggunakan 5 kode yang berbeda untuk proses data mining. Dengan preprocessing yang sama, saya menggunakan 3 metode klasifikasi yang berbeda. Yang pertama adalah Decision Tree (dilakukan sebanyak 2x), yang kedua dengan metode Naïve

Bayes (dilakukan sebanyak 2x), dan yang terakhir adalah dengan metode K-NN (dilakukan sebanyak 1x).

5. **Menentukan Algoritma yang Paling Cocok**

Dari hasil percobaan dengan berbagai macam metode dan kode program atau algoritma yang berbeda. Didapatkan hasil dari masing-masing percobaan yang saya dapatkan. Berikut hasilnya.

Hasil decision tree (1).

```
[ ] from sklearn.metrics import f1_score

    # Misalkan y_test adalah label yang benar dan y_pred adalah hasil prediksi
    f1_weighted = f1_score(y_test, y_pred, average='weighted')

    # f1_micro akan berisi skor F1 agregat
    print(f1_weighted)
```

```
0.7012522361359571
```

```
[ ] # accuracy
    from sklearn.metrics import accuracy_score

    accuracy_score(y_test, y_pred)
```

```
0.7836538461538461
```

```
[ ] # precision
    from sklearn.metrics import precision_score

    precision_score(y_test, y_pred, average='weighted')
```

```
0.6871794871794872
```

```
[ ] # recall
    from sklearn.metrics import recall_score

    recall_score(y_test, y_pred, average='weighted')
```

```
0.7836538461538461
```

Gambar 2. 1 Hasil Decision Tree Algoritma 1

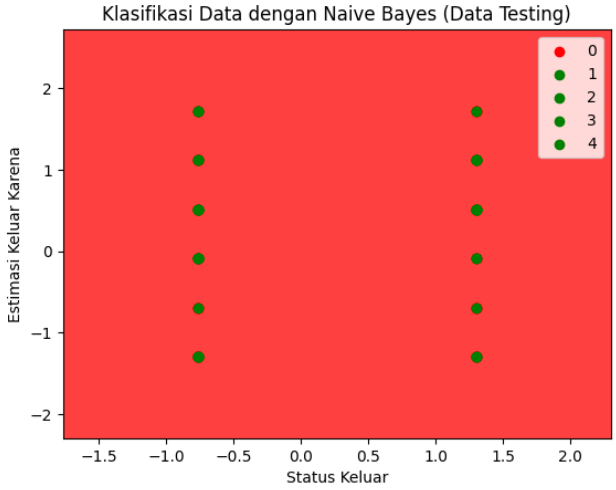
Hasil decision tree (2).

```
prediksiBenar = (hasilPrediksi == labelTesting).sum()
prediksiSalah = (hasilPrediksi != labelTesting).sum()
print("Prediksi Benar: ", prediksiBenar, "data")
print("Prediksi Salah: ", prediksiSalah, "data")
print("Akurasi: ", prediksiBenar/(prediksiBenar+prediksiSalah)* 100, "%")
```

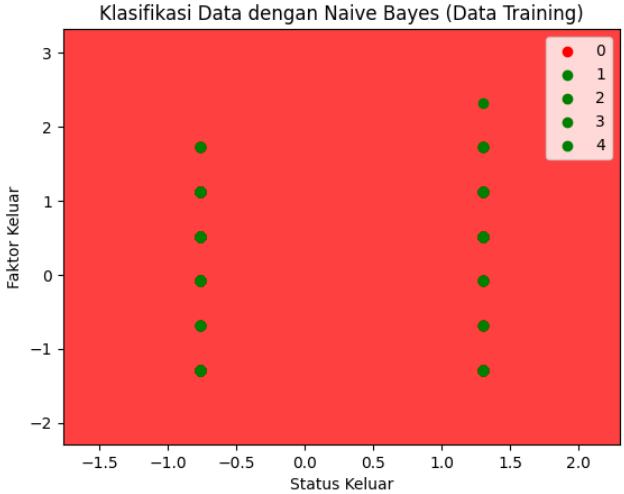
```
Prediksi Benar:  127 data
Prediksi Salah:  73 data
Akurasi:  63.5 %
```

Gambar 2. 2 Hasil Decision Tree Algorima 2

Hasil naïve bayes (1).

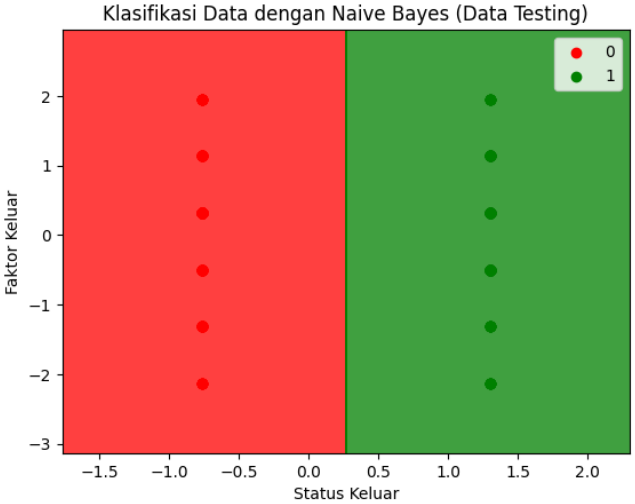


Gambar 2. 3 Hasil Naive Bayes Algortima 1 (Data Testing)

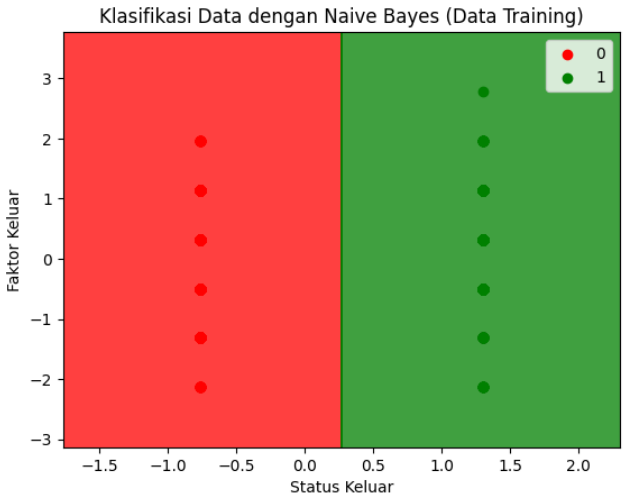


Gambar 2. 4 Hasil Naive Bayes Algorima 1 (Hasil Training)

Hasil naïve bayes (2)’

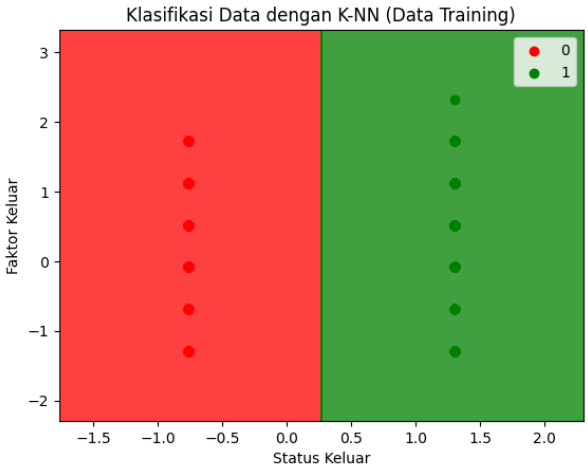


Gambar 2. 5 Hasil Naive Bayes Algoritma 2 (Hasil Testing)



Gambar 2. 6 Hasil Naive Bayes Algoritma 2 (Hasil Training)

Hasil K-NN (1).



Gambar 2. 7 Hasil K-Nearest Neighbor (KNN)

Dari kelima hasil diatas, didapatkan hasil untuk decision tree dengan akurasi 78% dan 63%. Walaupun mengguankan metode yang sama, namun terdapat perbedaan di dalam proses algoritmanya. Maka dari itu, lebih baik mengguankan decision tree 1 untuk melanjutkan tujuan dari proposal ini.

6. Eksperimen

Dalam melakukan eksperimen, penelitian ini menggunakan server komputer dengan processor AMD Ryzen 5 55000U with Radeon Graphics. Kapasitas memoory yaitu 8102MB RAM. Spesifikasi tersebut akan memudahkan dalam penelitian terutama untuk memproses ribuan citra digital. Penelitian ini menggunakan Bahasa Python dengan Jupiter Notebook sebagai aplikasi berbasis web yang digunakan untuk pengujian algoritma yang diusulkan.

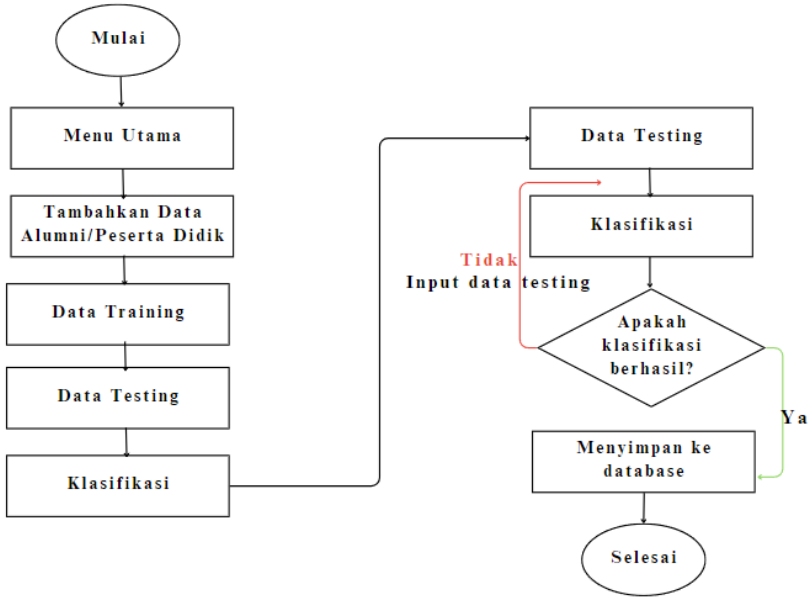
Pembagian dibagi menjadi 60% citra training, 20% citra validation, dan 20% citra testing. Citra training dan citra validation diguankan untuk melatih model deteksi sehingga dapat digunakan untuk mendeteksi citra testing dengan akurat.

7. Luaran

Tujuan akhir dari proses data mining ini adalah menghasilkan sebuah produk yang dapat digunakan oleh para guru dan staff yang bersangukutan untuk tujuan analisis dan prediksi siswa yang berpotensi untuk keluar karena kecocokan faktor dari hasil uji coba dataset training

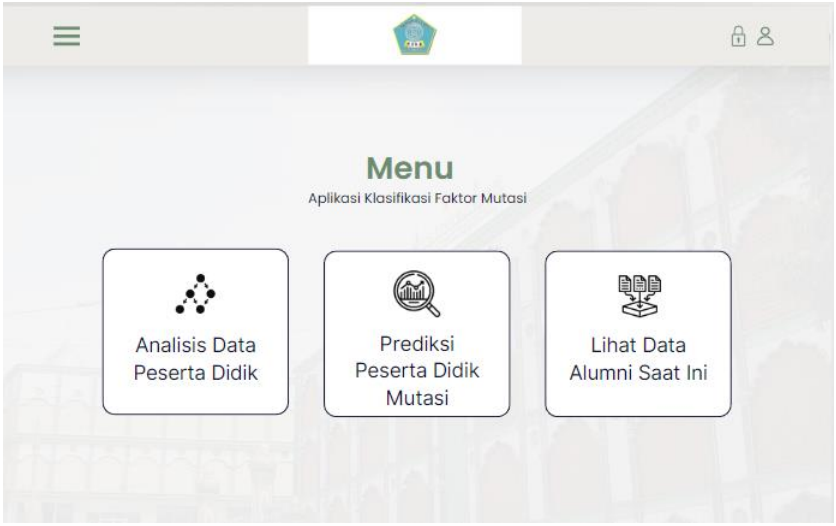
menggunakan decision tree. Produk yang akan dibangun oleh penulis adalah sebuah aplikasi berbasis web menggunakan sebuah framework django atau pilihan kedua dapat menggunakan steamlit.

Sebuah web aplikasi yang akan dibuat akan menampilkan hasil predikasi siswa yang berpotensi keluar. Data diambil dari data rekap data alumni, pada saat uji coba pertama, penulis akan mengguankan data mentahan yaitu dataset_alumni_SMK.csv untuk mengimpementasikan ke dalam web aplikasi. Cara kerja sistem web aplikasi adalah sebagai berikut.



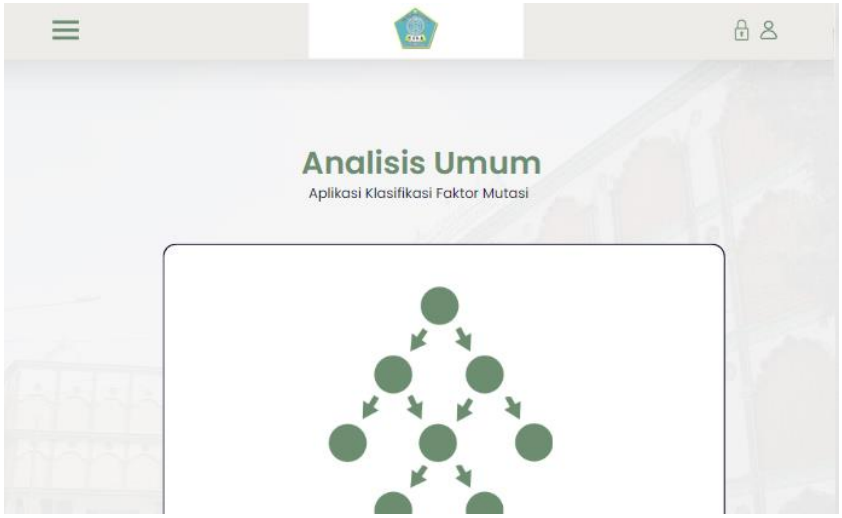
Gambar 3. 1 Flowchart Web Aplikasi (Luaran)

Setelah memahami flowchart dari data web aplikasi, dibuat rancangan design web yangn akan dibangun. Saat ini tampilan UI yang sudah dibuat oleh penulis proposal ini sebagai berikut.



Gambar 4. 1 UI Web Aplikasi

Gambar 4. 2 UI Web Aplikasi



Gambar 4. 3 UI Web Aplikasi

Gambar 4. 4 UI Web Aplikasi

JADWAL PENELITIAN

Jadwal penelitian disusun berdasarkan pelaksanaan penelitian, harap disesuaikan berdasarkan lama tahun pelaksanaan penelitian

Tabel 2. 1 Jadwal Penelitian

No.	Nama Kegiatan	Minggu Ke-												
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1.	Penyusunan dan pengajuan proposal		√												
2.	Pengumpulan dan analisis data		√												
3	FGD persiapan penelitian terkait permasalahan dan penyelesaiannya				√										
4	Pengembangan model						√	√	√						
5	Eksperimen							√	√						
6	Analisa hasil eksperimen								√						
7	Penyusunan publikasi jurnal									√	√				
8	Submit publikasi											√			
9	Penyusunan laporan kemajuan											√	√	√	
10	Penyusunan laporan akhir														√

DAFTAR PUSTAKA

Sitasi disusun dan ditulis berdasarkan sistem nomor sesuai dengan urutan pengutipan, mengikuti format Vancouver. Hanya pustaka yang disitasi pada usulan penelitian yang dicantumkan dalam Daftar Pustaka.

1. R. Liang, C. Huang, C. Zhang, B. Li, S. Saydam and I. Canbulat, "Exploring the Fusion Potentials of Data Visualization and Data Analytics in the Process of Mining Digitalization," in IEEE Access, vol. 11, pp. 40608-40628, 2023, doi: 10.1109/ACCESS.2023.3267813.
2. M. Li, H. Wang and J. Li, "Mining conditional functional dependency rules on big data," in Big Data Mining and Analytics, vol. 3, no. 1, pp. 68-84, March 2020, doi: 10.26599/BDMA.2019.9020019.
3. Mert Ozcan, Serhat Peker, A classification and regression tree algorithm for heart disease modeling and prediction, Healthcare Analytics, Volume 3, 2023, 100130, ISSN 2772-4425, https://doi.org/10.1016/j.health.2022.100130.
4. Ari Melo Mariano, Arthur Bandeira de Magalhães Lelis Ferreira, Maíra Rocha Santos, Mara Lucia Castilho, Anna Carla Freire Luna Campêlo Bastos, Decision trees for predicting dropout in Engineering Course students in Brazil, Procedia Computer Science, Volume 214, 2022, Pages 1113-1120, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2022.11.285.
5. Oluwaseun Adebayo A, Chaubey MS. Data Mining Classification Techniques on the Analysis of Student's Performance. GSJ: Volume 7, Issue 4. April 2019; ISSN 2320-9186. https://www.researchgate.net/publication/332233710
6. Khoirunnisa, Susanti L, Rokhmah IT, Stianingsih L. Prediksi Siswa SMK Al-Hidayah yang Masuk Perguruan Tinggi dengan Metode Klasifikasi. JURNAL INFORMATIKA. April 2021; 8(1): 26-33. ISSN: 2355-6579 | E-ISSN: 2528-2247. https://ejournal.bsi.ac.id/ejurnal/index.php/ji/article/view/9163