

**REKOMENDASI PROFIL LULUSAN DI SMK  
MENGUNAKAN METODE PCA DAN K - MEANS**

**IMELDA**

NRP. 2156102010

**DOSEN PEMBIMBING**

**Dr. Yohana Dewi Lulu Widyasari, S. Si, M. T.**

**Dr. Juni Nurma Sari, S.Kom, M.MT**

**PROGRAM STUDI  
MAGISTER TERAPAN TEKNIK KOMPUTER  
POLITEKNIK CALTEX RIAU  
2024**

HALAMAN PENGESAHAN

REKOMENDASI PROFIL LULUSAN DI SMK  
MENGUNAKAN METODE PCA DAN K-  
MEANS

**Imelda**

**NIM. 2156102010**

Tesis ini digunakan sebagai salah satu syarat untuk memperoleh  
Gelar Magister Terapan Teknik Komputer (M.Tr.Kom) di  
Politeknik Caltex Riau

Pekanbaru, 26 Februari 2024

Pembimbing  
Utama : Dr. Yohana Dewi Lulu W., S. Si, M. T.  
NIP. 007717

Pembimbing  
Pendamping : Dr. Jani Nurma Sari, S. Kom, M. MT  
NIP. 017218

Penguji : Dr. Dadang Syarif Sihabudin Sahid, S.Si.M.Sc.  
NIP. 007504

Penguji : Ananda, S.Kom., M.T., Ph.D.  
NIP. 108501

Penguji : Dini Normalasari, S.T., M.T.  
NIP. 048108

Disetujui oleh:

**Mengetahui,**  
Ketua Program Studi Magister Terapan Teknik Komputer  
Politeknik Caltex Riau



Dr. Umamah Nur Hafidha, S.T., M.Eng  
NIP. 0177501

## ABSTRAK

Penelitian ini mengeksplorasi dampak penyelenggaraan pendidikan vokasi, khususnya di SMK Negeri 6 Pekanbaru, terhadap jalur karir lulusan. Hasil dari *tracer study* tahun 2022 menunjukkan bahwa 24,69% lulusan memilih untuk kuliah, 36,63% bekerja, 31,28% masih mencari pekerjaan, dan 6,17% memutuskan berwirausaha. Implikasi temuan ini menyoroti perhatian terhadap kualitas pendidikan di SMK. Algoritma PCA dan K - Means digunakan untuk memberikan rekomendasi keberlanjutan karir lulusan.

Analisis yang dilakukan dengan PCA dan K-Means memberikan informasi bermanfaat tentang pola keputusan karir lulusan. Nilai *plot Silhouette* dengan integrasi PCA dengan K-Means di platform Orange naik lebih dari 0.430 dibandingkan dengan K-Means hanya 0,021. Untuk jurusan Rekayasa Perangkat Lunak dan Teknik Komputer Jaringan, mendominasi opsi bekerja dan menekankan betapa pentingnya menjalin kerjasama dengan dunia industri. Dengan memasukkan PCA, pemahaman tentang karakteristik lulusan menjadi lebih baik, dan ini memungkinkan rekomendasi yang lebih akurat.

Berdasarkan hasil penelitian dengan PCA dan K-Means, lulusan SMK Negeri 6 Pekanbaru menunjukkan bahwa 38% memilih untuk Bekerja, 32% memilih untuk Kuliah, 30% memilih untuk Bekerja dan Wirausaha dengan rekomendasi perlunya peningkatan dari fasilitas terkini yang belum optimal, dan adanya pembaharuan secara terus menerus dari info lowongan kerja yang dirasa masih kurang sehingga tujuan dari penelitian untuk memastikan bahwa pilihan sekolah di SMK sesuai dengan minat, bakat, dan potensi peserta didik, agar lulusan menjadi terampil dan berkualitas dapat tercapai. Implementasi sistem diharapkan mengurangi kesenjangan keterampilan dan kebutuhan dunia kerja, serta memberikan kontribusi positif pada manajemen pendidikan dan kualitas pembelajaran berbasis karakter di SMK.

Kata Kunci: Kualitas Belajar, SMK, *Unsupervised Learning*, PCA, K - Means

# DAFTAR ISI

<b>BAB I PENDAHULUAN</b> .....	<b>1</b>
<b>1.1. LATAR BELAKANG</b> .....	<b>1</b>
<b>1.2 PERMASALAHAN</b> .....	<b>3</b>
<b>1.3 TUJUAN</b> .....	<b>4</b>
<b>1.4 MANFAAT</b> .....	<b>4</b>
<b>1.5 SISTEMATIKA PENULISAN</b> .....	<b>5</b>
<b>BAB II KAJIAN PUSTAKA</b> .....	<b>6</b>
<b>2.1 TEORI PENUNJANG</b> .....	<b>6</b>
2.1.1. KUALITAS PEMBELAJARAN DI SMK .....	8
2.1.2. DEFINISI MACHINE LEARNING .....	9
2.1.3 K - MEANS.....	14
2.1.4 FEATURE RANK .....	16
2.1.5 PCA ( PRINCIPAL COMPONENT ANALYSIS ).....	18
2.1.6 FLASK .....	19
<b>2.2 PENELITIAN TERKAIT</b> .....	<b>20</b>
<b>BAB III DESAIN SISTEM</b> .....	<b>25</b>
<b>3.1 DESAIN SISTEM</b> .....	<b>26</b>
3.1.1 SUBJEK PENELITIAN.....	27
3.1.2 PROSEDUR PENELITIAN.....	27
<b>3.2 TAHAPAN MACHINE LEARNING</b> .....	<b>27</b>
3.2.1 IDENTIFIKASI MASALAH MACHINE LEARNING.....	28
3.2.2 DATA ACQUISITION .....	29
3.2.3 EXPLORATORY DATA ANALYSIS ( EDA ).....	33

3.2.4	PREPROCESSING .....	38
3.2.5	PCA .....	40
3.2.6	CLUSTERING .....	41
3.2.7	PERANCANGAN <i>DEPLOY WEB</i> .....	43
<b>BAB IV EKSPERIMEN DAN ANALISIS.....</b>		<b>49</b>
<b>4.1</b>	<b>PARAMETER EKSPERIMEN .....</b>	<b>49</b>
<b>4.2</b>	<b>KARAKTERISTIK DATA.....</b>	<b>49</b>
<b>4.3</b>	<b>SPEKIFIKASI PERALATAN UJI COBA .....</b>	<b>49</b>
<b>4.4</b>	<b>TAHAPAN MACHINE LEARNING PADA PYTHON.....</b>	<b>49</b>
4.4.1	DATA ACQUISITION .....	50
4.4.2	DATA EXPLORATION.....	50
4.4.3	<i>DATA PRE- PROCESSING</i> .....	51
4.4.4	MODELLING DENGAN PYTHON DENGAN FLASK .....	55
<b>4.5</b>	<b>EVALUASI MODELING K – MEANS DENGAN PYTHON.....</b>	<b>56</b>
<b>4.6</b>	<b>TAHAPAN MACHINE LEARNING DENGAN ORANGE .....</b>	<b>57</b>
4.6.1	DATA ACQUISITION DENGAN ORANGE .....	57
4.6.2	DATA EXPLORATION.....	58
4.6.3	MODELLING PADA ORANGE .....	59
<b>4.7</b>	<b>EVALUASI DENGAN ORANGE .....</b>	<b>61</b>
<b>4.8</b>	<b>PENGUJIAN SISTEM REKOMENDASI.....</b>	<b>63</b>
<b>BAB V KESIMPULAN DAN SARAN .....</b>		<b>69</b>
<b>5.1.</b>	<b>KESIMPULAN.....</b>	<b>69</b>
<b>5.2</b>	<b>SARAN .....</b>	<b>69</b>

## DAFTAR GAMBAR

Gambar 1.1 Pendidikan Yang Ditamatkan.....	2
Gambar 2.1 Tahapan Machine Learning.....	13
Gambar 2.2 Tahapan Pengolahan Data.....	14
Gambar 3.1 Flowchart Tahapan Penelitian.....	26
Gambar 3.2 Data Kuisisioner Rekomendasi.....	29
Gambar 3.3 Data Nilai .....	29
Gambar 3.4 Data Kuisisioner dan Nilai.....	30
Gambar 3.5 Informasi Data.....	33
Gambar 3.6 Informasi Jumlah Kolom.....	34
Gambar 3.7 Grafik Sebaran Berwirausaha.....	35
Gambar 3.8 Grafik Sebaran Alumni Kuliah.....	35
Gambar 3.9 Grafik Sebaran Alumni Bekerja.....	36
Gambar 3.10 Grafik Alumni Bekerja.....	36
Gambar 3.10 Heatmap .....	37
Gambar 3.11 Hasil Scatter Plot.....	38
Gambar 3.12 Cek Missing Values.....	39
Gambar 3.13 Normalisasi.....	40
Gambar 3.14 Penentuan Komponen PC.....	41
Gambar 3.15 Arsitektur Sistem.....	43
Gambar 3.16 Use Case Diagram.....	44
Gambar 3.17 <i>Activity Diagram</i> .....	45
Gambar 3.18 ERD.....	46
Gambar 3.19 Perancangan Halaman Depa.....	46
Gambar 3.20 Pemilihan Menu Pada Sistem.....	47
Gambar 3.21 Input Data Alumni.....	47

Gambar 3.22 Daftar Sample / Alumni .....	48
Gambar 4.1 Import Library di Python .....	50
Gambar 4.2 Memanggil Dari Dataset .....	50
Gambar 4.3 Cek <i>Missing Value</i> .....	51
Gambar 4.4 Jumlah Data Ganda .....	52
Gambar 4.5 Cek Data Type.....	52
Gambar 4.6 Cek Jumlah Data Points .....	53
Gambar 4.7 Deskripsi Dataset.....	53
Gambar 4.8 Korelasi .....	54
Gambar 4.9 Konfigurasi Flask .....	55
Gambar 4.10 Script K - Means dan PCA.....	55
Gambar 4.11 Evaluasi Pengujian Python.....	56
Gambar 4.12 Hasil rekomendasi .....	56
Gambar 4.13 <i>Widget File Orange</i> .....	57
Gambar 4.14 <i>Data Exploration</i> pada Orange .....	58
Gambar 4.15 K - Means dengan Orange.....	59
Gambar 4.16 Silhouette Plot K - Means .....	59
Gambar 4.17 PCA dan K - Means Orange.....	60
Gambar 4.18 Hasil K - Means dan PCA.....	60
Gambar 4.19 Hasil Scatter Plot.....	61
Gambar 4.20 Hasil PCA.....	62
Gambar 4.21 Halaman Login Pengujian.....	63
Gambar 4.22 Halaman Input Kuisisioner .....	63
Gambar 4.23 Data PCA dan K-Means.....	64
Gambar 4.24 Hasil Rekomendasi Sistem.....	64
Gambar 4.25 Hasil Menu Daftar Alumni.....	65
Gambar 4.26 Daftar Pertanyaan Kuisisioner .....	66
Gambar 4.27 Hasil Kuisisioner.....	66

## DAFTAR TABEL

Tabel 3.1 Atribut Dataset .....	32
Tabel 3.2 Penentuan Klaster .....	42
Tabel 4.1 Tabel Perbandingan Silhoutte .....	61
Tabel 4.2 Pengujian Blackbox Login.....	68
Tabel 4.3 Pengujian Input Alumni .....	68





# BAB I

## PENDAHULUAN

### 1.1. LATAR BELAKANG

SMK Bisa SMK Hebat dengan Sikap Kerja Santun Mandiri dan Kreatif merupakan motto yang diharapkan. Motto ini menunjukkan keahlian para peserta didik yang mampu mendorong generasi muda untuk terus berkreativitas dan menghasilkan karya-karya inovatif yang bermanfaat . Di sisi lain, SMK merupakan salah satu penerapan pendidikan vokasi yang dapat meningkatkan akses dan pemerataan maupun peningkatan kualitas. Pendidikan kejuruan melalui SMK mempersiapkan peserta didik terutama untuk bekerja pada bidang tertentu.

Secara garis besar, kualitas pendidikan SMK meliputi input, proses, output, hasil dan dampak. Input terdiri dari peserta didik baru, sumber daya manusia (SDM), kurikulum, infrastruktur sarana dan prasarana, keuangan, organisasi manajemen dan mitra; Proses tersebut meliputi penyesuaian kurikulum, pelaksanaan pembelajaran, penyelenggaraan *teaching factory* , pengembangan sumber daya manusia sekolah, penguatan kemitraan dengan dunia industri dan dunia kerja ( IDUKA ), penerapan budaya kerja, dan sertifikasi siswa. *Output* meliputi mutu lulusan sesuai Standar Kualifikasi Lulusan ( SKL ) dan persyaratan IDUKA serta sertifikasi akhir sesuai ( Kompetensi Kerja Nasional Indonesia ) Level II/III. Hasil merincikan keterserapan dan kemandirian lulusan, meningkatnya kemampuan lulusan dalam mengembangkan diri, serta berkembangnya kelembagaan SMK. Sementara, dampak terdiri dari meningkatnya kepercayaan masyarakat terhadap SMK dan terwujudnya sekolah mandiri.

Namun demikian, dalam realitas lapangan, pendidikan kejuruan di Sekolah Menengah Kejuruan ( SMK ) ternyata menduduki posisi kedua sebagai penyumbang terbesar tingkat pengangguran di Indonesia pada bulan Agustus 2022.

Pengangguran Terbuka Menurut Pendidikan Tertinggi yang Ditamatkan 1986 - 2022

No.	Pendidikan Tertinggi Yang Ditamatkan	2020		2021		2022	
		Februari	Agustus	Februari	Agustus	Februari	Agustus
1	Tidak/belum pernah sekolah	35,761	31,379	20,461	23,905	24,852	15,206
2	Tidak/belum tamat SD	346,778	428,813	342,734	431,329	437,819	663,125
3	SD	1,008,744	1,410,537	1,219,494	1,393,492	1,230,914	1,274,153
4	SLTP	1,251,352	1,621,518	1,515,089	1,604,448	1,460,221	1,500,807
5	SLTA Umum/SMU	1,748,834	2,662,444	2,305,093	2,472,859	2,251,558	2,478,173
6	SLTA Kejuruan/SMK	1,443,522	2,326,599	2,089,137	2,111,338	1,876,661	1,661,492
7	Akademi/Diploma	267,583	305,261	254,457	216,024	235,359	159,490
8	Universitas	824,912	981,203	999,543	848,657	884,769	673,485
	<b>Total</b>	<b>6,925,486</b>	<b>9,767,754</b>	<b>8,746,008</b>	<b>9,102,052</b>	<b>8,402,153</b>	<b>8,425,931</b>

Sumber: Survei Angkatan Kerja Nasional (Sakernas)

**Gambar 1.1 Pendidikan Yang Ditamatkan**

Sumber : (<https://www.bps.go.id>, 2022)

Pada Gambar 1.1 dapat dilihat data pengangguran terbuka menurut pendidikan tertinggi yang ditamatkan dari tahun 1986 – 2022, tamatan SMK pada bulan Agustus 2022 sekitar 1.661.492 orang. Kondisi tersebut juga dapat diamati di SMK Negeri 6 Pekanbaru. Data yang diperoleh dari Tracer Study pada tahun 2022 oleh Tim Humas terhadap para alumni menunjukkan bahwa 24,69% dari lulusan memilih untuk melanjutkan pendidikan ke jenjang kuliah, sementara 36,63% sudah memasuki dunia kerja. Di sisi lain, sekitar 31,28% lulusan belum memasuki dunia kerja, dan sisanya sebanyak 6,17% memilih untuk membuka usaha sendiri atau berwirausaha.

Penelitian dengan rekomendasi profil lulusan dari jurusan di bidang informatika, yaitu Rekayasa Perangkat dan Teknik Komputer Jaringan membantu penyelesaian permasalahan tingginya jumlah pengangguran dari tingkat pendidikan SMK yang dapat diidentifikasi dan diketahui cepat melalui rekomendasi profil lulusan yang akan bekerja, melanjutkan pendidikan atau berwirausaha. Langkah-langkah untuk memperkuat keterkaitan antara pendidikan, rekomendasi profil lulusan, dan kebutuhan pasar kerja dapat menjadi kunci untuk

menciptakan lulusan yang lebih siap dan relevan dengan tuntutan dunia kerja. Jika rekomendasi profil lulusan didasarkan pada kurikulum yang tidak selaras dengan kebutuhan aktual industri atau dunia kerja, lulusan mungkin tidak memiliki keterampilan yang diinginkan oleh pemberi kerja. Lulusan mungkin kesulitan menemukan pekerjaan yang sesuai dengan keterampilan yang dimiliki, sehingga terjadilah jumlah pengangguran makin tinggi.

Penelitian ini dilakukan dengan metode *machine learning* menggunakan algoritma K – Means dan PCA. Kelebihan dari *K – Means*, setiap titik data diberi label *cluster* yang sesuai, sehingga dapat memberikan pemahaman yang intuitif tentang kelompok-kelompok yang ada. Penelitian menggunakan data yang diambil dengan cara membagikan kuisioner kepada alumni tentang faktor yang berpengaruh terhadap kualitas pembelajaran di SMK seperti penyediaan fasilitas sarana prasarana, program kurikulum yang selaras dengan dunia industri dan dunia kerja. Selain itu, beberapa faktor dapat menjadi pemicu seperti: pemilihan jurusan dari peserta didik serta pembiayaan. Data tersebut akan diolah menggunakan K - Means untuk menghasilkan rekomendasi kepada peserta didik kelas XII untuk ke tahapan selanjutnya setelah tamat, yaitu bekerja, melanjutkan pendidikan, dan berwirausaha sehingga kemungkinan tingginya tingkat pengangguran dapat diketahui dan dihindari.

Hasil penelitian ini memberikan rekomendasi kepada sekolah agar lulusannya pada kompetensi keahlian di bidang informatika , yaitu Rekayasa Perangkat Lunak dan Teknik Komputer dan Jaringan dominan dapat bekerja, melanjutkan pendidikan, selebihnya berwirausaha adalah perlu adanya peningkatan kompetensi lebih lanjut terutama terhadap mata pelajaran selain Matematika, Bahasa Inggris, Dasar Kejuruan, Seni Budaya, Kompetensi dan PKN dalam pengembangan pendidikan berkarakter ( *soft skill* ) sehingga lulusan dapat berhasil melakukan pencapaian.

## **1.2 PERMASALAHAN**

Dari permasalahan tentang rekomendasi profil lulusan dengan tantangan seperti kurangnya informasi dan keberagaman dari peluang karir yang ditawarkan

SMK, kekhawatiran bahwa kurikulum SMK tidak selalu sejalan dengan kebutuhan industri atau dunia kerja. Melalui penelitian didapati bahwa upaya bagaimana mengubah dan mengatasi persepsi negatif bahwa SMK merupakan kontributor terbesar terhadap pengangguran di dunia pendidikan, walaupun kurikulum telah disesuaikan dengan kebutuhan masyarakat dan industri.

### **1.3 TUJUAN**

Tujuan penelitian ini adalah untuk memberikan rekomendasi profil lulusan SMK, dengan maksud untuk mengidentifikasi faktor-faktor yang mempengaruhi pilihan untuk bekerja, melanjutkan studi ke perguruan tinggi, atau memilih berwirausaha. Penelitian ini akan menggunakan pendekatan metode machine learning K-Means untuk mencapai tujuan tersebut.

### **1.4 MANFAAT**

1. Digunakan oleh peserta didik untuk memantapkan komitmennya untuk memanfaatkan peluang yang diberikan oleh pendidikan kejuruan.
2. Bagi Dinas Pendidikan Provinsi Riau bagian Perencanaan, menjadikan sebuah analisa pemikiran tentang rencana di tahun ini, untuk pembangunan satuan pendidikan baru yang disesuaikan dengan kebutuhan fasilitas.
3. Bagi Dunia Industri dan Dunia Kerja, dengan mengetahui rekoemendasi profil lulusan, dapat menyesuaikan penawaran pekerjaan dengan keterampilan, kecenderungan, dan potensi lulusan. Ini membantu menciptakan keselarasan antara kebutuhan perusahaan dan kemampuan individu dan juga mempercepat dan mempermudah proses rekrutmen dengan memberikan informasi yang lebih terperinci tentang keterampilan dan minat kandidat. Hal ini mengurangi waktu yang dibutuhkan untuk menemukan calon yang sesuai.

## **1.5 SISTEMATIKA PENULISAN**

Sistematika pembahasan dalam pengajuan tesis ini meliputi:

### **Bab 1 Pendahuluan**

Pada bagian pendahuluan ini memaparkan secara rinci tentang latar belakang dari penelitian, permasalahan, tujuan dan manfaat penelitian.

### **Bab 2 Kajian Pustaka**

Pada bagian Kajian Pustaka ini berisi teori – teori penunjang yang digunakan untuk penelitian ini dan menjelaskan penelitian – penelitian terdahulu yang pernah dilakukan yang terkait dengan penelitian ini dan menggambarkan tempat dilakukan penelitian ini.

### **Bab 3 Desain Sistem**

Pada bab ini menjelaskan secara detail desain sistem yang akan dikembangkan dimulai dari tahapan-tahapan .

### **Bab 4 Jadwal dan Anggaran**

Pada bab ini menjelaskan secara detail jadwal dan anggaran yang dibutuhkan

## **BAB II**

### **KAJIAN PUSTAKA**

Badan Pusat Statistik pada tahun 2022 melaporkan bahwa pengangguran di Indonesia menurut jenjang pendidikan dari tingkat SMK menempati posisi kedua terbanyak yaitu sebesar 9.42%, bersumber dari : [Badan Pusat Statistik \(bps.go.id\)](https://bps.go.id). Lulusan SMK ingin langsung bekerja, tetapi tidak terserap di dunia usaha. Hal ini disebabkan oleh meningkatnya lulusan SMK yang tidak diimbangi oleh kesempatan kerja, sumber dari : [Pengangguran di Indonesia Paling Banyak Lulusan SMK \(katadata.co.id\)](https://katadata.co.id). Kondisi tersebut terjadi karena dipengaruhi oleh beberapa faktor penentu kualitas pembelajaran di SMK, yang tidak hanya menghasilkan lulusan yang langsung bekerja, tetapi juga melanjutkan pendidikan ke jenjang yang lebih tinggi

#### **2.1 TEORI PENUNJANG**

Pendidikan menurut Jandhyala BG Tilak (2002:46) dalam Buku Panduan Profesional Riset Pendidikan Asia-Pasifik sebagai berikut: pendidikan umum atau pelatihan kejuruan? Ini adalah pilihan di banyak negara berkembang. Dalam kehidupan manusia, pendidikan umum menciptakan “orang biasa” dan pendidikan profesi dan kejuruan menciptakan “orang istimewa”. Pelatihan kejuruan memiliki keunggulan dalam memberikan keterampilan spesifik pekerjaan yang dapat meningkatkan persiapan pekerja untuk pekerjaan yang mereka temukan dan membuat mereka lebih produktif.

Prosser dan Quigley menyatakan dalam Surya Dharma (2013) bahwa inti dari pendidikan kejuruan adalah mengajarkan cara berpikir dan bekerja melalui pelatihan yang berulang-ulang. Ada tiga cara yang harus diajarkan, yaitu:

- Kebiasaan beradaptasi dengan lingkungan kerja,
- Kebiasaan dalam proses pelaksanaan pekerjaan dan
- Kebiasaan pikiran (di tempat kerja).

Ketika seseorang masuk ke dunia kerja, penguasaan *hard skill* cenderung statis, kecuali mereka dapat beradaptasi dengan perkembangan baru yang terkait dengan *hard skill*. Namun, penguasaan *soft skill* bersentuhan dan bersinggungan langsung dalam kegiatan sehari-hari. Penguasaan *soft skill* sangat penting dalam perkembangan usia peserta didik karena *soft skill* lah yang akan membimbing lebih jauh karakter peserta didik ketika yang bersangkutan masuk ke dalam dunia industri, usaha, dunia kerja . Menghadapi situasi ini, penguatan *soft skill* atau pendidikan karakter sangat penting. Pengertian, tahapan, dan manfaat pendidikan karakter termasuk dalam materi pendidikan karakter. Sebagai contoh, nilai-nilai kesopanan mencakup pengertian kesopanan, tindakan untuk menjadi sopan, dan manfaatnya. Ada kebutuhan untuk mengembangkan metode pembelajaran yang mampu untuk membangun karakter peserta didik yang memiliki keterampilan *soft* dan *hard*, dan kesemuanya itu tertuang pada mata pelajaran yang menggali pendidikan karakter seperti Pendidikan Agama, Pendidikan Kewarganegaraan, Sejarah dan lain – lain.

Wenrich dan Galloway (1988) berpendapat bahwa pelatihan kejuruan sama dengan pelatihan teknis dan pelatihan kejuruan. Istilah pelatihan kejuruan, pelatihan spesialis, dan pelatihan kejuruan dapat digunakan secara bergantian. Istilah-istilah ini memiliki arti yang berbeda, tetapi ketiga istilah tersebut adalah pelatihan untuk bekerja.

Selain itu, Wenrich dan Galloway berpendapat bahwa pendidikan profesional dapat diartikan sebagai pendidikan khusus yang mempersiapkan siswa untuk pindah ke pekerjaan tertentu, tugas keluarga, atau untuk memperkuat keterampilan tenaga kerja. Calhoun (1982) menyatakan bahwa pendidikan kejuruan adalah program pendidikan yang mempersiapkan individu siswa menjadi tenaga profesional yang juga ingin melanjutkan pendidikan ke jenjang yang lebih tinggi. Clarke dan Winch (2007) mendefinisikan VET sebagai pendidikan yang mempersiapkan kaum muda dan orang dewasa untuk memasuki kehidupan profesional, di mana pembelajaran berorientasi pada masalah dan praktik.

Henry dan Thomson dalam Berg (2002) menjelaskan VET ( Vocational Education and Training ) sebagai berikut, VET setara dengan belajar “melakukan



pekerjaan”, VET bertujuan untuk meningkatkan keterampilan teknis dan posisi seseorang di lingkungan melalui manajemen teknologi, dan VET terkait erat dengannya. terhadap kebutuhan pasar. Oleh karena itu, seringkali dilihat sebagai faktor yang mempengaruhi perekonomian nasional.

Dari sudut pandang para ahli dapat disimpulkan bahwa VET merupakan pendidikan yang bertujuan untuk menghasilkan sumber daya manusia yang kompeten, cakap, dan produktif dalam bidang tertentu. Pelatihan vokasi diselenggarakan dengan membawa “suasana kerja ke dalam sekolah” agar siswa belajar untuk “bekerja”. Pelatihan vokasi membekali peserta didik dengan pengetahuan, keterampilan dan pengalaman untuk bekerja secara efektif dan efisien, serta kemampuan untuk senantiasa beradaptasi dan mengantisipasi perkembangan ilmu pengetahuan dan teknologi.

### **2.1.1. KUALITAS PEMBELAJARAN DI SMK**

Menurut (Mukhidin Mukhidin, 2019) , ada hubungan antara kualitas sekolah dan kondisi infrastruktur sekolah. *Self-efficacy* siswa memiliki pengaruh yang kuat terhadap keputusan pendaftaran siswa SMA. Pemerintah harus meningkatkan anggaran infrastruktur untuk sekolah kejuruan (SMK), membantu sekolah untuk memobilisasi dana pengembangan modal serta membantu memerangi publisitas negatif yang mempengaruhi sekolah kejuruan.

Kualitas sekolah dipengaruhi oleh beberapa faktor, seperti kualitas guru, kualitas peserta didik, ukuran kelas, lingkungan yang baik dan bersih, ruang kelas yang berventilasi baik, perpustakaan yang lengkap, pembelajaran online. Ukuran kelas mempengaruhi partisipasi peserta didik di kelas. Beberapa penelitian yang menguji pengaruh kualitas sekolah terhadap, misalnya, ukuran kelas, karakteristik guru dan hasil belajar peserta didik, dengan menggunakan ukuran per kapita, menghasilkan kesimpulan yang berbeda.

Kualitas infrastruktur memiliki kaitan penting dengan kualitas mutu sekolah dan penentu dari sekolah berkualitas sehingga dengan pendidikan berbasis sekolah saja tidak menghasilkan lulusan yang berkompeten tinggi jika

dibandingkan dengan sistem ganda atau magang di dunia industri, usaha ataupun dunia kerja karena sebagian besar sekolah kejuruan kekurangan peralatan dan mesin yang relevan dan dalam beberapa situasi ekstrim, mereka menggunakan mesin dan peralatan yang sudah ketinggalan zaman.

Penelitian berjudul 'Rekomendasi Profil Lulusan di SMK menggunakan Metode Machine Learning K-Means dan PCA dengan Feature Rank' diharapkan mampu memberikan solusi terhadap pemilihan program kejuruan di SMK yang sesuai dengan minat individu, dengan tujuan agar lulusan dapat menghadapi pilihan karir yang lebih sesuai, seperti bekerja, melanjutkan pendidikan, atau berwirausaha.

### **2.1.2. DEFINISI MACHINE LEARNING**

Menurut Brownlee, J. (2017). *Machine Learning Mastery* *Machine learning* merupakan sebuah algoritma. Machine learning adalah cabang ilmu komputer yang berfokus pada pengembangan teknik dan algoritma yang memungkinkan komputer untuk belajar dan mengambil keputusan atau melakukan prediksi berdasarkan data tanpa harus secara eksplisit diprogram. Dalam machine learning, komputer mampu "belajar" dari data yang ada dan mengenali pola atau hubungan yang tersembunyi dalam data tersebut. Proses pembelajaran ini dilakukan dengan memanfaatkan algoritma dan model matematika yang dikembangkan secara khusus.

Ada beberapa pendekatan dalam machine learning, termasuk *supervised learning* (pembelajaran terbimbing), *unsupervised learning* (pembelajaran tanpa pengawasan), dan *reinforcement learning* (pembelajaran penguatan). Dalam *supervised learning*, model dilatih menggunakan data yang telah diberi label atau target, dengan tujuan untuk melakukan prediksi atau klasifikasi pada data baru. Dalam *unsupervised learning*, model mencoba mengenali pola atau struktur dalam data yang tidak memiliki label atau target sebelumnya. Sedangkan dalam *reinforcement learning*, model belajar melalui interaksi dengan lingkungan dan mendapatkan umpan balik berdasarkan keputusan yang diambil.

*Machine learning* memiliki berbagai aplikasi yang luas, termasuk pengenalan wajah, pengenalan ucapan, deteksi penipuan, analisis sentimen, pengoptimalan bisnis, prediksi keuangan, dan masih banyak lagi. Berbagai industri dan bidang semakin menaruh perhatian pada machine learning karena perkembangan teknologi dan ketersediaan data yang semakin meningkat. Dalam *machine learning*, terdapat dua jenis utama algoritma: *supervised learning* dan *unsupervised learning*. Berikut adalah penjelasan singkat tentang kedua jenis algoritma tersebut:

#### A. *Supervised Learning*

Dalam *supervised learning*, model dilatih menggunakan data yang telah diberi label atau target. Data latihan terdiri dari pasangan input dan output yang sudah diketahui. Tujuan dari algoritma *supervised learning* adalah untuk membuat model yang dapat mempelajari hubungan antara *input* dan *output*, sehingga dapat memprediksi output yang benar untuk input baru yang belum pernah dilihat sebelumnya. Beberapa metode dalam *supervised learning*, di antaranya:

1. **Klasifikasi (Classification):** Model melakukan prediksi pada kelas atau label yang telah ditentukan. Contohnya, memprediksi apakah sebuah email adalah spam atau bukan.
2. **Regresi (Regression):** Model melakukan prediksi pada nilai kontinu. Contohnya, memprediksi harga rumah berdasarkan fitur-fitur tertentu.

Beberapa algoritma supervised learning yang populer meliputi: Decision Trees, Random Forests, Support Vector Machines (SVM), Naive Bayes, dan Neural Networks.

#### B. *Unsupervised Learning*

Dalam *unsupervised learning*, model memproses data yang tidak memiliki label atau target sebelumnya. Tujuan dari algoritma *unsupervised learning* adalah untuk menemukan pola atau struktur tersembunyi dalam data tanpa bimbingan eksternal. Beberapa metode dalam unsupervised learning, di antaranya:

1. **Clustering:** Model mengelompokkan data ke dalam kelompok-kelompok yang memiliki kesamaan berdasarkan fitur-fitur yang ada.

2. Deteksi anomali ( *Anomaly detection* ): Tugas deteksi anomali adalah mengidentifikasi data yang berbeda atau tidak biasa dalam suatu dataset. Tujuannya adalah untuk menemukan data yang tidak mengikuti pola umum atau perilaku yang diharapkan.
3. Reduksi dimensi ( *Dimensionality reduction* ): Tugas reduksi dimensi adalah mengurangi jumlah atribut atau fitur dalam dataset. Tujuannya adalah untuk menghilangkan atribut yang tidak relevan atau redundan dan mewakili data dengan dimensi yang lebih rendah namun mempertahankan informasi yang penting bahkan menemukan representasi yang lebih ringkas.
4. Pembelajaran aturan *asosiasi* ( *Association rule learning* ): Tugas pembelajaran aturan *asosiasi* adalah menemukan hubungan atau keterkaitan antara item atau atribut dalam dataset. Tujuannya adalah untuk mengidentifikasi pola atau aturan yang sering muncul bersama-sama dalam data.

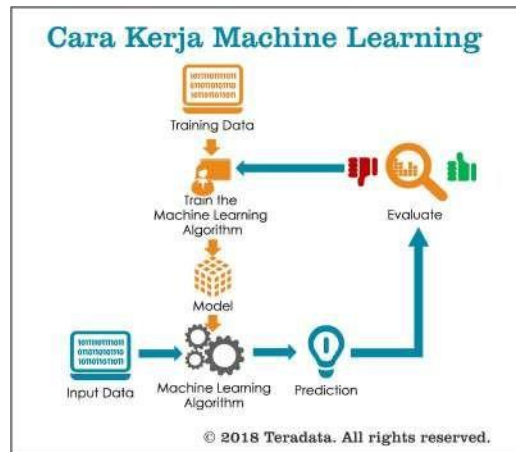
Beberapa algoritma unsupervised learning yang umum digunakan adalah: K-Means, Hierarchical Clustering, Principal Component Analysis (PCA), dan t-SNE.

Kedua jenis algoritma ini memiliki peran penting dalam *machine learning* dan digunakan untuk memecahkan berbagai jenis masalah dengan pendekatan yang berbeda. Pemilihan algoritma yang tepat tergantung pada jenis data, tugas yang ingin diselesaikan, serta karakteristik masalah yang dihadapi. Tahapan dalam machine learning melibatkan beberapa langkah penting untuk membangun dan mengimplementasikan model yang efektif. Berikut adalah tahapan umum dalam proses machine learning:

1. Pemahaman Masalah: Langkah pertama adalah memahami masalah atau pertanyaan yang ingin dijawab melalui machine learning. Identifikasi tujuan dan memahami data yang tersedia.
2. Pengumpulan Data: Kumpulkan data yang relevan untuk masalah yang ingin diselesaikan. Data ini bisa berupa dataset yang sudah ada atau perlu dikumpulkan melalui sumber-sumber yang sesuai.

3. Pembersihan dan Pemrosesan Data: Lakukan pembersihan data dengan menghapus nilai yang hilang atau tidak valid, menangani pencilan (*outliers*), dan melakukan transformasi data yang diperlukan. Proses ini juga melibatkan pemilihan fitur yang relevan untuk analisis lebih lanjut.
4. Pembagian Data: Bagi data menjadi set pelatihan ( *training set* ), set validasi (validation set), dan set pengujian ( *testing set* ) . *Training set* digunakan untuk melatih model, set validasi digunakan untuk memilih model terbaik dan mengoptimalkan parameter, sedangkan set pengujian digunakan untuk menguji performa model yang final.
5. Pemilihan Model: Pilih model machine learning yang sesuai dengan jenis masalah yang ingin diselesaikan. Misalnya, regresi linier untuk masalah prediksi nilai numerik, klasifikasi biner untuk masalah klasifikasi dua kelas, atau jaringan saraf tiruan untuk tugas-tugas yang lebih kompleks.
6. Pelatihan Model: Latih model menggunakan set pelatihan yang telah dibagi sebelumnya. Model akan belajar dari data dan menyesuaikan parameter-nya untuk meminimalkan kesalahan atau memaksimalkan kinerja sesuai tujuan yang ditentukan.
7. Evaluasi Model: Evaluasi performa model menggunakan set validasi. Metrik evaluasi yang umum digunakan termasuk akurasi, presisi, recall, F1-score, atau metrik khusus lainnya tergantung pada jenis masalah yang dihadapi.
8. Optimasi Model: Lakukan optimasi model dengan mencoba berbagai konfigurasi, mengubah parameter, atau menggunakan teknik seperti penyetelan hiperparameter (hyperparameter tuning) untuk meningkatkan performa model.
9. Pengujian Model: Gunakan set pengujian yang belum pernah dilihat oleh model sebelumnya untuk menguji performa model yang sudah dioptimasi. Hal ini membantu memastikan bahwa model dapat menggeneralisasi dengan baik ke data baru.

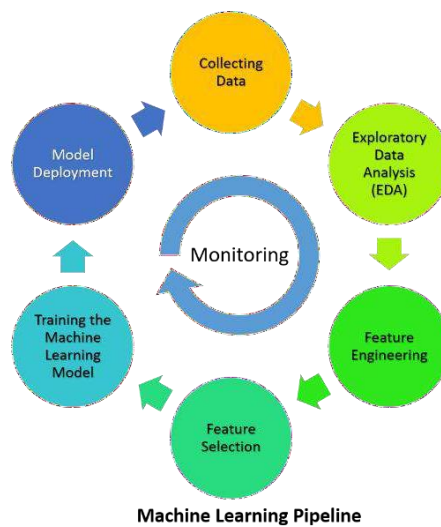
10. Penyampaian dan Implementasi Model: Setelah model dianggap memadai, model tersebut dapat diimplementasikan dalam lingkungan produksi atau digunakan untuk membuat prediksi pada data baru.



Gambar 2.1 Tahapan Machine Learning

Sumber dari : <https://www.advernesia.com/wp-content/uploads/2018/05/cara-kerja-machine-learning.png>

Pada Gambar 2.1, untuk menghasilkan suatu model, komputer akan melakukan proses belajar (training) dari data tersebut. Proses ini menggunakan algoritma machine learning, yang menggunakan teknik statistika. Model ini menghasilkan informasi yang dapat digunakan sebagai pengetahuan untuk memecahkan masalah sebagai proses input-output. Selain itu, model ini memiliki kemampuan untuk melakukan klasifikasi dan prediksi. Data akan dibagi menjadi kumpulan data pembelajaran ( kumpulan data pembelajaran ) dan kumpulan data pengujian ( kumpulan data pengujian ). Cara pembagian data ini berbeda - beda tergantung pada algoritma yang digunakan.



Gambar 2.2 Tahapan Pengolahan Data

Sumber : <https://dqqlab.id/tahapan-algoritma-data-science-simpel-cocok-untuk-pemula>

Pada Gambar 2.2, tahapan dalam membuat algoritma pengolahan data yang menggabungkan *Data Science* dan *Machine Learning* dimulai dari mendefinisikan data, pengumpulan dan persiapan data, EDA. Exploratory Data Analysis (EDA) adalah suatu proses uji investigasi awal yang bertujuan untuk mengidentifikasi pola, menemukan anomali, menguji hipotesis dan memeriksa asumsi. Di dalamnya juga termasuk proses penanganan missing values, outlier, reduksi dimensi, pengelompokkan, transformasi dan distribusi data dan penempatan nilai random, bersumber dari : <https://klc2.kemenkeu.go.id/kms/knowledge/exploratory-data-analysis-eda-dengan-orange-data->. Selanjutnya, *feature engineering*, *feature selection*, *training the learning machine model*, dan *model deployment*.

### 2.1.3 K - MEANS

K - Means adalah salah satu algoritma klusterisasi yang populer dan mudah dipahami. Menurut (Gustientiedina, 2019) , *Clustering* merupakan bidang penelitian dalam analisis dan data mining. Pada *clustering* partisi dari objek data yang mempunyai karakteristik sama akan dikelompokkan pada satu kelompok dan

data yang memiliki karakteristik berbeda akan dikelompokkan pada kelompok yang lainnya. Salah satu metode pada *clustering* yang terkenal daripada algoritma *clustering* lainnya adalah K-means karena kesederhanaan algoritma dan efisiensinya Tujuan utama dari algoritma K-Means adalah membagi sekumpulan data menjadi beberapa kelompok atau klaster berdasarkan kesamaan atribut. Algoritma ini mengasumsikan bahwa setiap data akan dikelompokkan ke dalam salah satu klaster yang memiliki mean (rata-rata) yang paling dekat dengan data tersebut. Berikut adalah langkah-langkah utama dalam algoritma K-means:

1. *Inisialisasi*: Tentukan jumlah klaster yang diinginkan (K) sebelumnya. Langkah pertama adalah menginisialisasi K titik pusat klaster secara acak. Titik-titik ini disebut centroid dan akan menjadi representasi awal setiap klaster.
2. *Assignment*: Setiap data akan diatribusikan ke klaster yang memiliki centroid terdekat dengannya berdasarkan jarak Euclidean atau metrik jarak lainnya. Setiap data akan menjadi anggota dari satu klaster tertentu berdasarkan jarak terkecil dengan centroid.
3. *Update Centroid*: Setelah semua data ditugaskan ke klaster, langkah selanjutnya adalah memperbarui posisi centroid. Posisi centroid diupdate dengan menghitung rata-rata dari semua data yang termasuk dalam klaster tersebut. Ini akan menghasilkan centroid baru yang lebih akurat.
4. *Iterasi*: Langkah 2 dan 3 diulang secara berulang sampai tercapai kondisi berhenti. Kondisi berhenti dapat berupa konvergensi, yaitu saat tidak ada perubahan yang signifikan dalam penugasan data ke klaster atau perubahan posisi centroid. Selain itu, jumlah iterasi maksimum yang ditentukan juga dapat menjadi kondisi berhenti.
5. *Output*: Setelah algoritma konvergen, outputnya adalah klasterisasi final, di mana setiap data telah diatribusikan ke salah satu klaster. Hasil klasterisasi ini dapat digunakan untuk analisis lebih lanjut atau pengambilan keputusan.
6. K - Means memiliki beberapa kelebihan, seperti kecepatan komputasi yang cepat dan mudah diimplementasikan. Namun, algoritma ini juga



memiliki beberapa kelemahan, seperti sensitivitas terhadap inisialisasi centroid awal, ketergantungan pada jumlah kluster yang ditentukan sebelumnya, dan ketidakmampuannya untuk menangani data dengan bentuk kluster yang kompleks atau ukuran kluster yang tidak merata.

7. Selain itu, algoritma K-means juga memiliki variasi, seperti K - Means++ yang memperbaiki inisialisasi centroid awal untuk menghindari hasil yang suboptimal, dan K-means dengan jumlah kluster yang dapat berubah (misalnya, *K-means* dengan *Hierarchical K-means*).

#### 2.1.4 FEATURE RANK

*Feature Ranking* adalah proses pengurutan fitur-fitur dalam dataset berdasarkan tingkat kepentingan atau informativitasnya terhadap tugas yang ingin diselesaikan. Ini membantu untuk mengidentifikasi fitur-fitur yang paling relevan atau memiliki dampak signifikan dalam memprediksi atau menjelaskan variabel target. Kegunaan utama dari feature ranking adalah sebagai berikut:

- A. Seleksi Fitur (*Feature Selection*): *Feature ranking* membantu dalam memilih fitur-fitur yang paling informatif atau relevan, sehingga dapat mengurangi dimensi data dan meningkatkan kinerja model. Dengan mengurangi jumlah fitur, waktu pelatihan model dapat berkurang, overfitting dapat dicegah, dan interpretabilitas model dapat ditingkatkan.
- B. Pemahaman Data: Dengan menganalisis peringkat fitur, dapat dipahami hubungan antara fitur-fitur dan target yang ingin diprediksi. Ini membantu dalam menemukan wawasan tentang data dan memvalidasi atau mengoreksi asumsi tentang pentingnya fitur-fitur tertentu.
- C. Efisiensi Komputasional: Dengan mengurutkan fitur-fitur berdasarkan peringkatnya, dapat diprioritaskan penggunaan fitur-fitur yang paling penting dalam model. Hal ini dapat mengurangi waktu komputasi yang diperlukan untuk melatih dan menerapkan model.
- D. Interpretabilitas Model: Dalam beberapa kasus, menggunakan hanya subset fitur-fitur yang paling penting dapat meningkatkan interpretabilitas

model. Dengan menghilangkan fitur-fitur yang kurang penting atau redundan, dapat dibuat model lebih mudah dipahami oleh manusia dan menjelaskan hubungan antara fitur-fitur dan hasil prediksi.

Ada beberapa metode yang dapat digunakan untuk melakukan peringkat fitur. Berikut ini adalah beberapa di antaranya:

1. Metode *Filter*:

Metode *filter* menggunakan statistik atau metrik yang tidak terkait dengan algoritma pembelajaran tertentu untuk mengukur kepentingan fitur secara independen. Contoh metode filter termasuk:

- a. *Korelasi*: Mengukur hubungan linier antara fitur dan variabel target.
- b. *Chi-square*: Digunakan untuk fitur kategorikal dan mengukur hubungan antara fitur dan variabel target yang bersifat kategorikal.
- c. *Information Gain*: Mengukur perbedaan informasi sebelum dan setelah pemisahan berdasarkan fitur dalam K – Means.
- d. Metode *Wrapper*:

Metode *wrapper* menggunakan algoritma pembelajaran untuk mengevaluasi setiap kombinasi fitur dan memilih subset fitur yang memberikan performa terbaik. Contoh metode wrapper termasuk:

1. *Recursive Feature Elimination (RFE)*: Menghapus secara iteratif fitur-fitur yang kurang penting berdasarkan performa model.
2. *Forward Selection*: Membangun model dengan satu fitur pada setiap iterasi dan menambahkan fitur dengan performa terbaik

e. Metode *Embedded*:

1. Metode *embedded* mengevaluasi kepentingan fitur secara langsung selama proses pembelajaran model. Contoh metode *embedded* termasuk:

2. *Regularization* (misalnya, *Lasso* dan *Ridge regression*): Memperkenalkan penalti untuk koefisien fitur, mengurangi bobot fitur yang kurang penting.
3. *K - Means Feature Importance*: Menggunakan atribut seperti *Sillhouette Score* dalam K - Means untuk menentukan peringkat fitur.

f. Metode Dimensionality Reduction:

Metode *dimensionality reduction*, seperti *Principal Component Analysis* (PCA) dan t-SNE, dapat mengurangi dimensi fitur dengan mempertahankan sebagian besar informasi penting dalam data.

### 2.1.5 PCA ( PRINCIPAL COMPONENT ANALYSIS )

Principal Component Analysis (PCA) digunakan sebagai metode statistik dan matematika yang digunakan untuk mengurangi dimensi data. Tujuan utama PCA adalah untuk menemukan pola-pola yang ada dalam data dengan mengubahnya ke dalam ruang dimensi yang lebih rendah. Dengan melakukan ini, PCA membantu menyederhanakan kompleksitas data sambil tetap mempertahankan informasi penting. Menurut (Marukatat, 2023) , *Principal Component Analysis (PCA) is one of the most widely used data analysis methods in machine learning and AI. This manuscript focuses on the mathematical foundation of classical PCA and its application to a small-sample-size scenario and a large dataset in a high-dimensional space scenario.* Menurut (Kurita, 2019), *Principal component analysis (PCA) is a standard tool in modern data analysis and is used by almost all scientific disciplines. The goal of PCA is to identify the most meaningful basis to reexpress a given data set. It is expected that this new basis will reveal hidden structure in the data set and filter out the noise. There are so many applications such as dimensionality reduction, data compression, feature extraction, and data visualization.*

Proses PCA melibatkan transformasi linier dari ruang dimensi tinggi ke ruang dimensi rendah; dimensi yang lebih rendah ini disusun berdasarkan urutan pentingnya untuk menjelaskan variasi data. Komponen utama pertama, atau

komponen utama, adalah arah data dengan variasi maksimum, dan komponen utama kedua adalah arah dengan variasi maksimum yang tersisa, dan seterusnya. Menurut (Nasution, 2019) , PCA berfungsi sebagai model yang kuat untuk menganalisis data.

Analisis data multivariat, pengurangan dimensi, dan pemahaman struktur data yang menggunakan PCA adalah aplikasi yang umum. Menurut (Retno, 2023), Metode Principal Component Analysis atau PCA merupakan suatu teknik multivariat yang bertujuan untuk mereduksi faktor atau variabel dalam jumlah besar menjadi beberapa faktor yang lebih sedikit. Selain digunakan untuk mereduksi jumlah variabel, metode PCA juga dapat digunakan untuk menangani masalah multikolinearitas dengan mereduksi jumlah variabelnya. Analisis PCA dapat digunakan untuk mereduksi dimensi suatu data tanpa mengurangi karakteristik data tersebut secara signifikan. (Putra, 2016). PCA dapat mereduksi dimensi pada data besar tanpa menghilangkan seluruh informasi sehingga dapat meningkatkan kualitas hasil pengelompokan k-means clustering. (Yulianti, 2022)

### **2.1.6 FLASK**

Flask adalah sebuah framework web mikro yang ditulis dalam bahasa pemrograman Python. Dibuat pada tahun 2010 oleh Armin Ronacher, Flask dimaksudkan untuk menjadi ringan dan mudah digunakan sehingga mempercepat pengembangan aplikasi web. Beberapa fitur utama Flask adalah sebagai berikut:

- **Mikroframework:** Flask dirancang sebagai mikroframework, yang berarti ia memiliki fungsionalitas dasar yang diperlukan untuk membuat aplikasi web, tetapi juga sangat fleksibel dalam hal mengembangkan dan menggunakan modul tambahan sesuai kebutuhan.
- **Ruting:** Mekanisme routing digunakan oleh Flask untuk menghubungkan URL ke fungsi tertentu dalam kode aplikasi. Ini memudahkan pengembang untuk menentukan bagaimana aplikasi menanggapi permintaan HTTP.

## 2.2 PENELITIAN TERKAIT

Penelitian ini merupakan studi literatur yang merujuk beberapa penelitian terdahulu, di antaranya sebagai berikut:

“74% merasa jenuh belajar di SMK seperti mata pelajaran social, yang pembelajaran prakteknya terbatas disebabkan karena keterbatasan fasilitas, guru dan dukungan industry”. (Suharno, Pambudi and Harjanto, 2020)

" Pemahaman kondisi sekolah berdasarkan pandangan siswa dan lulusan perlu dianalisis, terutama untuk melakukan proses pengendalian kondisi di lapangan. Studi ini juga menangkap pandangan para wisudawan mengenai kegiatan mereka setelah lulus” (1. Ariyani LF & 8.)

“Ditemukan bahwa, ada hubungan antara kualitas sekolah dan kondisi infrastruktur sekolah. Self-efficacy siswa memiliki pengaruh yang kuat terhadap keputusan pendaftaran siswa SMA. Pemerintah harus meningkatkan anggaran infrastruktur (Mukhidin Mukhidin, 2019)

Penelitian berjudul “Rekomendasi Profil Lulusan di SMK menggunakan PCA dan K – Means” didapat dari acuan beberapa penelitian terdahulu dan memperkuat penelitian dengan judul di atas nantinya dengan menerapkan machine learning dan perbandingan antara 2 algoritma.

Adapun keterbaruan penelitian dibandingkan dengan penelitian terdahulu adalah perbandingan teknik PCA dan K - Means dengan *feature rank* .

Berdasarkan penelitian yang dilakukan “K – Means merupakan salah satu algoritma yang dapat digunakan untuk prediksi dan dapat menggunakan data kategorikal maupun data numerikal. K-Means merupakan salah satu metode data clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster* atau kelompok. Kelebihan algoritma K-Means diantaranya adalah mampu mengelompokkan objek besar dan pencilan obyek dengan sangat cepat sehingga mempercepat proses pengelompokan. Penelitian menggunakan analisa K – Means yang melihat bagaimana penggunaan metode tersebut dalam pengelompokkan penyebaran penyakit menular pada manusia berdasarkan set variabel yang dibuat untuk setiap puskesmas di tiap kecamatan, yang berjumlah 32 di Kabupaten Majalengka. Enam di antaranya adalah penyakit menular seksual, diare, ISPA, kusta, malaria, tuberkulosis, dan penyakit menular seksual. Kelebihan penelitian, metode pengolahan data menggunakan algoritma K-Means dapat digunakan untuk mengelola persediaan di Puskesmas Pandanaran yang memungkinkan petugas untuk memiliki daftar Puskesmas dengan pasien

terbanyak yang menderita penyakit menular sehingga apabila persediaan obat akan dibeli pada dapat diketahui secara cepat. Kekurangannya, K – Means hasil pengelompokan tidak unik (selalu berubah), dan prosesnya cepat, tetapi tidak akurat.” BASTIAN, A. (2018) ‘Penerapan Algoritma K-Means Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka)’, *Jurnal Sistem Informasi*, 14(1), pp. 28–34. doi: 10.21609/jsi.v14i1.566.

Penelitian lain dari Nasari, F. and Darma, S. (2015) ‘Seminar Nasional Teknologi Informasi dan Multimedia 2015 Penerapan K-MEANS Clustering Pada Data Penerimaan Mahasiswa Baru (Studi Kasus: Universitas Potensi Utama) pp. 6–8.’ mengenai K - Means adalah “ Penelitian menggunakan algoritma K – Means *clustering* untuk mengelompokkan data pada kelompok yang sama, dan data yang berbeda pada kelompok yang berbeda, sehingga pengelompokkan data Penerimaan mahasiswa baru Universitas Potensi Utama lebih terstruktur. dengan hasil klaster pertama terdiri dari jurusan Sistem Informasi jika sekolahnya adalah SMA, dan jurusan Teknik Informatika jika sekolahnya adalah SMK. Kelebihannya, pada tahap ini penulis mengaplikasikan data pendaftaran mahasiswa baru menggunakan algoritma K-Means Clustering pada Aplikasi Matlab 6.1, untuk 2 kelompok data. Kelompok pertama terlihat pada titiktitik penyebaran yang ditandai dengan warna hijau dan pusat cendroid pertama terletak pada titik 2.907216495; 1.654639175; 1.474226804 yang ditandai dengan kotakhitam disudut gambar. Kelompok kedua ditandai dengan titik-titik penyebaran warna merah dengan pusat clusternya pada titik 1 ; 1.75; 1.5 dan ditandai dengan kotak hitam sebelah tengah gambar. Kekurangan, perlu mengkaji lebih dalam dan teliti dalam penentuan nilai cendroid awal yang digunakan dan jumlah data yang digunakan, karena perbedaan pengambilan data pusat cendroid awal, mempengaruhi hasil cluster yang digunakan juga akan mempengaruhi hasil cendroid akhirnya.”

Berdasarkan penelitian Asroni, A., Fitri, H. and Prasetyo, E. (2018) ‘Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokkan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik)’, *Semesta Teknika*, 21(1), pp. 60–64. doi: 10.18196/st.211211.,

Menurut penulis, algoritma K – Means digunakan penelitian untuk mengetahui kesenjangan jumlah mahasiswa baru yang diterima terhadap jumlah pendaftar serta

pengelompokkan data calon mahasiswa baru di Universitas Muhammadiyah Yogyakarta dengan studi kasus di Fakultas Kesehatan dan Ilmu Keperawatan dan Fakultas Ilmu Sosial dan Ilmu Politik. Dengan diketahui nilai kesenjangan tersebut akan menentukan kebijakan pimpinan dalam melakukan promosi pada masing-masing Fakultas. Kelebihan metode dengan K – Means, terjadi pengulangan tahapan sehingga terdapat kestabilan, dan diuji dengan aplikasi WEKA sehingga dapat ditentukan bahwa pendidikan dokter dan ilmu hubungan internasional menjadi jurusan pilihan calon mahasiswa Universitas Muhammadiyah Yogyakarta. Untuk pengujian, diperlukan kapasitas penyimpanan sehingga dapat menggunakan aplikasi WEKA yang menampung data lebih banyak.

Metode *k-means cluster analysis* bisa menjadi solusi untuk pengklasifikasian. Penelitian lain Surohman, Surohman, Fabrianto, Luky Riza, Faiza Faizah, Novianti M (2021) ‘ Korelasi Antara Profil dan Nilai Akademis Siswa dengan Menggunakan Algoritma K-Means’. doi : 10.25126/jtiik.2021843034 mengenai Pengelompokan siswa berdasarkan aspek sosial dengan pendekatan klasterisasi k-Means dapat digunakan. Menurut penulis, ada bentuk atau pola korelasi antara profil pribadi siswa terhadap prestasi akademis yang diraih oleh siswa dengan membuat klasterisasi dari data yang ada menggunakan algoritma K-Means, sehingga membantu para pendidik untuk mencari solusi yang berkaitan dengan peningkatan nilai akademis masing-masing siswa. Kelebihannya, ada 36% atau 184 dari total 512 siswa yang letak klaster profil dan nilai akademisnya bersesuaian. 43 siswa dengan nilai rendah dan ekonomi keluarga yang minim dan 83 siswa dengan nilai akademis yang tertinggi serta kondisi ekonomi keluarga yang berkecukupan. Kekurangannya, pendekatan algoritma atau metode lain dapat digunakan, nilai rapor juga dapat digunakan untuk penelitian serupa, dan akan lebih baik jika orangtua mengisi profil data dengan lengkap dan sesuai dengan pilihan yang ada.

Berdasarkan penelitian (Azizah, 2023) Algoritma k-means dapat digunakan dalam kasus segmentasi pelanggan kartu kredit untuk mengelompokkan pelanggan kartu kredit berdasarkan faktor-faktor seperti jumlah transaksi, jumlah pembayaran, riwayat kredit, dan lain-lain. Dengan demikian, perusahaan dapat memahami kebutuhan dan preferensi pelanggan kartu kredit dengan lebih baik, sehingga dapat menyediakan barang dan jasa yang lebih sesuai dengan kebutuhan pelanggan. Melalui *K-Means Clustering* yang diperjelas dengan analisis PCA, pengelola mall dapat memahami pengunjungnya dengan lebih baik agar dapat mengambil keputusan pemasaran dan pemasangan produk yang tepat. Menurut parameter gender, wanita lebih sering mengunjungi mall. Adapun menurut usianya, pengunjung berusia 30an merupakan target pemasaran yang tepat dan harus dimaksimalkan memenuhi kebutuhannya agar didapatkan hasil yang maksimal. Kelebihannya, dengan PCA, dimudahkan visualisasi hasil *cluster* agar lebih mudah dipahami. Terdapat 3 *cluster* yang terbentuk. Untuk pengujian, membutuhkan kapasitas penyimpanan data yang berlebih dalam menggunakan aplikasi *Rapidminer*.

Menurut (Fikri, 2021) , pengelompokan semua Kabupaten dan Kota di Indonesia berdasarkan kedekatan nilai dari berbagai variabel kemiskinan seperti jumlah penduduk, pendapatan perkapita, angka melek huruf dan akses pendidikan, akses air bersih dan layak, dan faktor lain dapat menyebabkan perbedaan ini yang dirilis oleh BPS pada tahun 2020 sehingga dapat menunjukkan bagaimana lembaga sosial swasta dan pemerintah dapat menggunakan data yang dihasilkan dari pengelompokan / tersebut untuk membuat keputusan dan kebijakan terbaik untuk mengatasi kemiskinan di Indonesia. Selain itu, dapat berfungsi sebagai referensi bagi dunia akademik dan pendidikan untuk mengembangkan penelitian yang mencakup lingkup data dan bidang lainnya. Kelebihannya, hasil dari analisis kelompok dengan metode K-Means, yang menggunakan variabel PCA 0, PCA 1, dan PCA 2, serta jumlah klaster sebanyak 4, menunjukkan bahwa jumlah anggota kelompok Kabupaten/Kota dalam kelompok 1 adalah 117, kelompok 2 adalah 154, kelompok 3 adalah 173, dan kelompok 4 adalah 70. Dilakukan interpretasi terhadap keempat cluster yang terbentuk, dan ditemukan bahwa klaster 4 berada pada level 4, yang berarti bahwa cluster ini membutuhkan perhatian khusus untuk



mengatasi kemiskinan. Dan sektor utama yang diperlukan dalam pengentasan kemiskinan adalah pendidikan, karena di klaster 4 tingkat tamatan pendidikan dan angka melek baca memiliki rata-rata terendah karena rendahnya sektor pendidikan berdampak pula terhadap tingginya angka pengangguran. Kapasitas penyimpanan data juga harus disediakan berlebih, karena aplikasi KNIME juga diperlukan untuk pengujian.

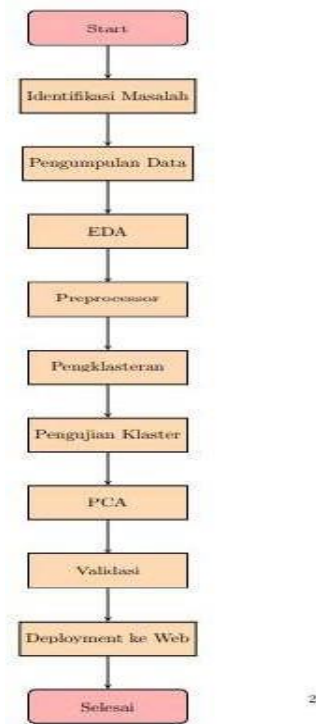
Menurut (Dana, 2021) , Penelitian yang menggunakan *machine learning* untuk menganalisa pengelompokan peserta pada uji kompetensi pada Lembaga Sertifikasi Profesi ( LSP ) yang dibagi menjadi 3 klaster pada *Network Security Monitoring* pada skema Desain Database Web dengan algoritma PCA dan K – Means yaitu 38% Kompeten pada klaster 1, 32 % pada klaster 2 dengan hasil Kompeten pada beberapa bagian dari *Network Security Monitoring* dan 30% dari klaster 3 dinyatakan tidak kompeten. Kelebihannya, dengan menggunakan PCA pada algoritma K-Means, proses pengelompokan dapat dioptimalkan untuk menentukan titik centroid awal algoritma K-Means. Tiga klaster yang ideal terbentuk dari pengelompokan data yang dihasilkan dari uji kompetensi yang dilakukan di lembaga sertifikasi profesional, masing-masing disebut sebagai "Klaster 1", "Klaster 2", dan "Klaster 3". Kekurangannya, uji kompetensi "Desain Database untuk Web" hanya dilakukan pada penilaian yang berasal dari Program Studi Teknik Informatika pada semester 6. Para peserta harus memiliki pemahaman yang kuat tentang pengembangan aplikasi web. Para mahasiswa pada program Studi Akuntansi Komputer dan Manajemen Informasi harus lebih mahir dalam materi Pengembangan Aplikasi Web melalui kegiatan pengayaan.

## **BAB III**

### **DESAIN SISTEM**

Bab ini menjelaskan tentang desain sistem, data penelitian, analisis data, analisa kebutuhan kerja, skenario pengujian, tempat dan jadwal penelitian, dan rencana anggaran biaya. Bagian data penelitian meliputi informasi data yang digunakan dan metode pengumpulan data, analisis data meliputi Exploratory Data Analysis (EDA). Bagian skenario pengujian meliputi proses dari beberapa metode pengujian yang digunakan pada sistem. Bagian jadwal pelaksanaan berisi rentang waktu pelaksanaan penelitian.

### 3.1 DESAIN SISTEM



Gambar 3.1 Flowchart Tahapan Penelitian

Tahapan penelitian yang akan dilakukan penyusun dalam proses penelitian:

Pada gambar 3.1, terdapat flowchart tahapan penelitian yaitu:

1. Melakukan identifikasi masalah studi literatur dengan mempelajari dan memahami teori-teori yang digunakan, yaitu mengenai prediksi dan perangkaian, unsupervised learning dan K - Means
2. Menentukan dan mengumpulkan data – data yang dibutuhkan dalam penelitian yaitu data nilai dan hasil kuisioner seperti : Apakah sekarang saudara saat ini seorang siswa SMK atau sudah lulus.
3. Melakukan EDA ( Exploratory Data Analysis )

4. Melakukan perancangan Machine Learning seperti :
  - a. Preprocessing data dengan normalisasi
  - b. Missing values
5. Melakukan pengklasteran
6. Melakukan implementasi dan pengujian dengan memodelkan data dengan metode K - Means.
7. Menganalisa hasil dari pengujian dengan menggunakan metode di atas dan menentukan faktor yang paling berpengaruh dalam hal rekomendasi output lulusan dengan menggunakan PCA ( *Principal Component Analysis* ).
8. Melakukan proses *deployment* ke dalam bentuk web dengan menggunakan Python.
9. Perancangan Dashboard

### **3.1.1 SUBJEK PENELITIAN**

Subjek penelitian terdiri dari alumni yang menamatkan pendidikan dari berbagai Angkatan SMK Negeri 6 Penelitian yang dipilih dengan menggunakan secara acak melalui form yang dibagikan kepada wali kelas

### **3.1.2 PROSEDUR PENELITIAN**

Data dikumpulkan melalui Google Form yaitu melalui survei online. Partisipan diberikan informasi tentang tujuan penelitian dan diminta memberikan persetujuan sebelum mereka terlibat.

### **3.2 TAHAPAN MACHINE LEARNING**

Tahapan Machine Learning dimulai dari proses identifikasi masalah, pengumpulan data, penyiapan data, pemodelan, uji coba pemodelan, evaluasi pemodelan, peningkatan pemodelan dan menentukan prediksi.

Dataset berupa *Data Collection* yang berasal dari Data nilai selama menjadi peserta didik di SMK dan data yang berasal dari hasil jawaban kuisioner yang dibagikan kepada peserta didik dan alumni sebanyak 30 Atribut.

### 3.2.1 IDENTIFIKASI MASALAH MACHINE LEARNING

Penyelenggaraan pendidikan vokasi di SMK Negeri 6 Pekanbaru mendefinisikan beberapa masalah sehingga diterapkan *machine learning K – Means* dan *PCA* sebagai berikut :

1. Kesenjangan antara ketrampilan yang diajarkan dan kebutuhan dunia kerja. Implementasi sistem machine learning menggunakan algoritma K-Means dan PCA akan membantu mengurangi kesenjangan antara keterampilan yang diajarkan di SMK dan kebutuhan dunia kerja. Hal ini mengisyaratkan bahwa ada kebutuhan untuk menyesuaikan kurikulum dan metode pembelajaran agar lebih sesuai dengan tuntutan industri dan dunia kerja.
2. Rendahnya Pengetahuan dan Pemahaman Minat, Bakat, dan Potensi Peserta Didik  
Implementasi penerapan *machine learning* memberikan rekomendasi kepada peserta didik berdasarkan klaster tertentu. Ini menunjukkan bahwa mungkin terdapat kurangnya pemahaman tentang minat, bakat, dan potensi siswa dalam pengambilan keputusan karir. Perlu dipertimbangkan untuk meningkatkan metode penilaian atau bimbingan karir.
3. Kualitas pembelajaran di SMK  
Terdapat upaya untuk menilai kualitas pembelajaran di SMK, termasuk data nilai peserta didik, ketersediaan sarana dan prasarana, dan program kurikulum yang selaras dengan dunia kerja dan industri. Ini menunjukkan bahwa ada perlunya peningkatan elemen-elemen ini untuk menjamin kualitas pendidikan yang lebih baik.

### 3.2.2 DATA ACQUISITION

The image shows a screenshot of a Google Form titled "Data Kuisiioner Rekomendasi". The form contains numerous questions related to student preferences and perceptions, such as "Apakah Saudara saat ini sebagai siswa SMK/SMK Pekanbaru atau sudah lulus?", "Apakah Saudara pernah atau sedang mengikuti atau akan mengikuti pendidikan di SMK/SMK?", and "Apakah Saudara akan melanjutkan ke jenjang pendidikan yang lebih tinggi?". Below the questions, there are several rows of responses from different students, including their names, IDs, and answers to the various questions.

Gambar 3.2 Data Kuisiioner Rekomendasi

Pada gambar 3.2, Data kuisiioner didapat dari hasil survey yang dibagikan kepada alumni melalui Google Form.

Pada gambar 3.3, Data nilai didapat dari rekap nilai ijazah pada Bagian Kurikulum SMK Negeri 6 Pekanbaru.

Jurusan : RPL															
NO	NAMA SISWA	NO.PESERTA UJIAN	PAI	PKN	B.INDO	MTK	SEJARAH	B.ING	SENBUK	PJOK	SIMDIG	FISIKA	KIMIA	D.PROG - KEJURU AN	KOMPE TENSI
1	ALIF RAFA SHIDIQ	09-01-218-159	85	85	80	85	70	80	75	75	70	70	70	82	82
2	DHEA PUSPITA SARI	09-01-218-160	92	96	84	96	98	98	90	98	89	89	97	97	86
3	DIAN ADRINI YULIADI	09-01-218-161	96	96	98	96	89	90	96	88	98	97	93	95	89
4	DWI ANDINI	09-01-218-162	92	94	96	97	93	92	94	95	95	98	91	95	89
5	FADHIL HABIBI	09-01-218-163	95	98	88	89	91	98	98	98	96	96	97	96	89
6	FADLI	09-01-218-164	89	94	89	96	97	90	92	96	96	96	95	95	88
7	FARHAN AZIZ FIANDRA	09-01-218-165	94	95	88	97	93	88	96	96	97	96	98	93	87
8	IVAN ANANDA	09-01-218-166	94	92	88	90	85	91	98	92	97	95	96	91	85
9	JAYADI KARYADI	09-01-218-167	95	92	82	97	91	96	86	80	89	95	98	92	87
10	JEFRI SETIAWAN	09-01-218-168	96	97	95	96	91	94	98	98	98	96	95	97	89
11	KHALID SALMAN MARPAUNG	09-01-218-169	90	94	88	86	80	89	86	84	80	85	80	87	83
12	KIKI WAHYU SAPUTRA	09-01-218-170	88	94	82	85	83	87	85	82	85	80	80	85	83
13	LAILANI DESPI YENTI	09-01-218-171	95	93	86	87	91	87	96	90	85	98	91	95	87
14	LIDYA DAMAYANTI	09-01-218-172	97	98	98	97	89	94	88	98	97	97	95	95	88
15	M. FARIS DARMAWAN	09-01-218-173	96	97	97	85	91	98	98	96	96	96	96	95	94
16	M. RIZAL	09-01-218-174	94	98	80	96	91	85	97	97	86	85	83	90	82

Gambar 3.3 Data Nilai

Gambar 3.4 Data Kuisioner dan Nilai

Pada gambar 3.4 Dataset Kuisioner didapat dengan membagikan kuisioner melalui *Google Form* dengan berbagai macam pertanyaan. Selanjutnya, dataset dalam penelitian ini menggunakan dataset dari penggabungan data hasil respon kuisioner alumni dan data nilai selama di SMK Negeri 6 Pekanbaru dengan beberapa atribut Hasil penggabungannya dapat dilihat pada Tabel 3.1.

No.	Atribut	Tipe Data
1.	Jurusan Anda di SMK 6 Pekanbaru?	Numerik
2.	Jenis Kelamin	Categorial
3.	Anda Berdomisili Dimana?	Categorial
4.	Usia anda saat ini?	Numerik
5.	Apakah Anda merupakan anak pertama?	Categorial
6.	Apakah Ayah Anda pernah atau sedang menjalankan sebuah usaha/ bisnis saat Anda SMK?	Categorial
7.	Pendidikan terakhir Ayah?	Numerik
8.	Apakah Ibu Saudara pernah atau sedang menjalankan sebuah usaha/ bisnis saat Anda SMK?	Categorial
9.	Pendidikan terakhir Ibu?	Numerik
10.	Apakah Saudara pernah atau sedang	Categorial

	menjalankan sebuah usaha/ bisnis saat masih sekolah SMK?	
11.	Apakah Saudara aktif dalam organisasi Sekolah/ekstrakurikuler?	Categorial
12.	Apakah Saudara pernah mengikuti sebuah kursus saat sekolah di SMK?	Categorial
13.	SMK tempat Saudara belajar berstatus sebagai	Categorial
14.	Apakah alasan masuk SMK?	Numerik
15.	SMK tempat Saudara belajar menyediakan fasilitas yang lengkap?	Numerik
16.	Apakah SMK tempat Saudara belajar menyediakan fasilitas yang up to date (terkini)	Categorial
17.	Bagaimanakah biaya sekolah di SMK menurut Saudara saat masih bersekolah disana?	Numerik
18.	Apakah Saudara pernah mendapatkan beasiswa saat di SMK?	Categorial
19.	Apakah Saudara melanjutkan ke perguruan tinggi setelah lulus SMK?	Categorial
20.	Jika pertanyaan di atas adalah "Ya", pada jenjang pendidikan apakah Anda melanjutkan ?	Numerik
21.	Jika ada yang bertanya, apakah Saudara akan merekomendasikan kepada keluarga/ teman/ orang lain untuk bersekolah di SMK tempat Saudara pernah belajar?	Categorial



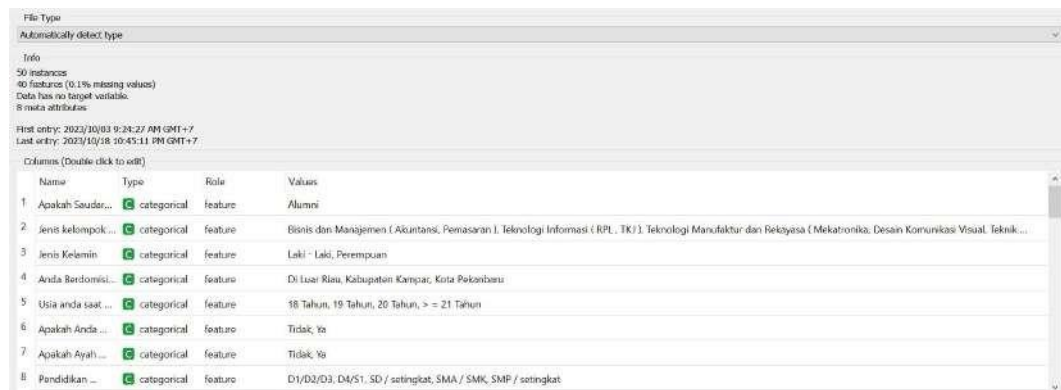
22.	( Pertanyaan untuk Alumni ) Apakah Saudara pernah atau sedang menjalankan sebuah bisnis setelah lulus SMK?	Categorial
23.	( Pertanyaan untuk Alumni ) Apakah Saudara Bekerja saat ini?	Categorial
24.	( Pertanyaan untuk Alumni ) Apakah Bidang Pekerjaan Saudara saat ini sesuai dengan jurusan pada saat SMK?	Categorial
25.	( Pertanyaan untuk Alumni ) Apakah SMK membantu Saudara dalam mendapatkan pekerjaan?	Categorial
26.	PAI	Numerik
27.	PKN	Numerik
28.	B.INDO	Numerik
29.	MTK	Numerik
30.	SEJARAH	Numerik
31.	B.ING	Numerik
32.	SENBUD	Numerik
33.	PJOK	Numerik
34.	SIMDIG	Numerik
35.	FISIKA	Numerik
36.	KIMA	Numerik
37.	D.PROG. KEJURUAN	Numerik
38.	KOMPETENSI	Numerik

Tabel 3.1 Atribut Dataset

### 3.2.3 EXPLORATORY DATA ANALYSIS ( EDA )

Salah satu bentuk dalam melakukan analisis data adalah dengan melakukan *Exploratory Data Analysis* ( EDA ). Pada EDA, dilakukan eksplorasi data sehingga akan mendapatkan informasi yang tepat terkait dataset. EDA merupakan suatu proses untuk mempelajari bentuk data yang akan menentukan bagaimana proses pengolahan terhadap data tersebut. Proses *data exploration* merupakan proses mengeksplorasi dataset, yaitu melihat isi dataset, apakah mempunyai *missing value*, kemudian melakukan transformasi dari data kategori menjadi data numerik

#### A. Informasi Data



The screenshot shows a software interface for data analysis. At the top, it says 'File Type' and 'Automatically detect type'. Below that, it provides 'Info' including '50 instances', '40 features (0.1% missing values)', 'data has no target variable', and '8 missing attributes'. It also shows the 'First entry' and 'Last entry' dates. The main part of the screenshot is a table titled 'Columns (Double click to edit)' with the following data:

	Name	Type	Role	Values
1	Apakah Saudara...	☑ categorical	feature	Alumni
2	Jenis kelompok ...	☑ categorical	feature	Bisnis dan Manajemen ( Akuntansi, Pemasaran ), Teknologi Informasi ( RPL, TKJ ), Teknologi Manufaktur dan Rekayasa ( Mekatronika, Desain Komunikasi Visual, Teknik...
3	Jenis Kelamin	☑ categorical	feature	Laki - Laki, Perempuan
4	Anda Berdomisi...	☑ categorical	feature	Di Luar Riau, Kabupaten Kampar, Kota Pekanbaru
5	Usia anda saat ...	☑ categorical	feature	18 Tahun, 19 Tahun, 20 Tahun, > = 21 Tahun
6	Apakah Anda ...	☑ categorical	feature	Tidak, Ya
7	Apakah Ayah ...	☑ categorical	feature	Tidak, Ya
8	Pendidikan ...	☑ categorical	feature	D1/D2/D3, DA/S1, SD / setingkat, SMA / SMK, SMP / setingkat

Gambar 3.5 Informasi Data

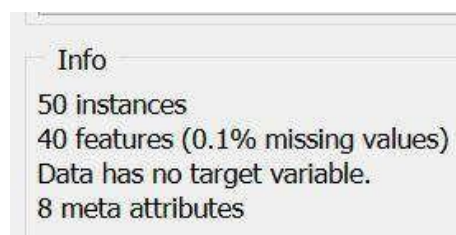
Pada gambar 3.5, dapat dilihat bahwa Data Kuisioner Rekomendasi dan Data Nilai memiliki 50 *Instances* dan 40 *Features* dan kebanyakan bertipe data *categorical*.

#### B. Informasi Kolom

Melihat informasi data type dari setiap kolom. Digunakan untuk mengidentifikasi awal kolom apa saja yang akan digunakan. Data didapat dari kuisioner yang dibagikan kepada alumni melalui *Google Form*, terdiri dari 40 atribut dengan tipe data kategori, di antaranya : Jenis Kelompok SMK? , Apakah Anda merupakan anak pertama?, Apakah Ayah Anda pernah atau sedang

menjalankan sebuah usaha/ bisnis saat Anda SMK?, Pendidikan terakhir Ayah?, Apakah Ibu Saudara pernah atau sedang menjalankan sebuah usaha/ bisnis saat Anda SMK?, Pendidikan terakhir Ibu? , Apakah Saudara pernah atau sedang menjalankan sebuah usaha/ bisnis saat masih sekolah SMK?, Apakah Saudara aktif dalam organisasi Sekolah/ekstrakurikuler?, Apakah Saudara pernah mengikuti sebuah kursus saat sekolah di SMK?, SMK tempat Saudara belajar menyediakan fasilitas yang lengkap?, Apakah SMK tempat Saudara belajar menyediakan fasilitas yang up to date (terkini), Bagaimanakah biaya sekolah di SMK menurut Saudara saat masih bersekolah disana? Apakah Saudara pernah mendapatkan beasiswa saat di SMK?, Apakah Saudara melanjutkan ke perguruan tinggi setelah lulus SMK?, Jika pertanyaan nomor 21 adalah "Ya", pada jenjang apakah Anda melanjutkan ?, Jika ada yang bertanya, apakah Saudara akan merekomendasikan kepada keluarga/ teman/ orang lain untuk bersekolah di SMK tempat Saudara pernah belajar?, Apakah materi yang diajarkan di sekolah sesuai dengan kebutuhan dunia kerja Saudara?, Apakah Saudara pernah atau sedang menjalankan sebuah bisnis setelah lulus SMK? Apakah Saudara pernah atau sedang menjalankan sebuah bisnis setelah lulus SMK?, Apakah Saudara Bekerja saat ini?, Apakah Bidang Pekerjaan Saudara saat ini sesuai dengan jurusan pada saat SMK?, Apakah SMK membantu Saudara dalam mendapatkan pekerjaan?

### C. Informasi Jumlah Data

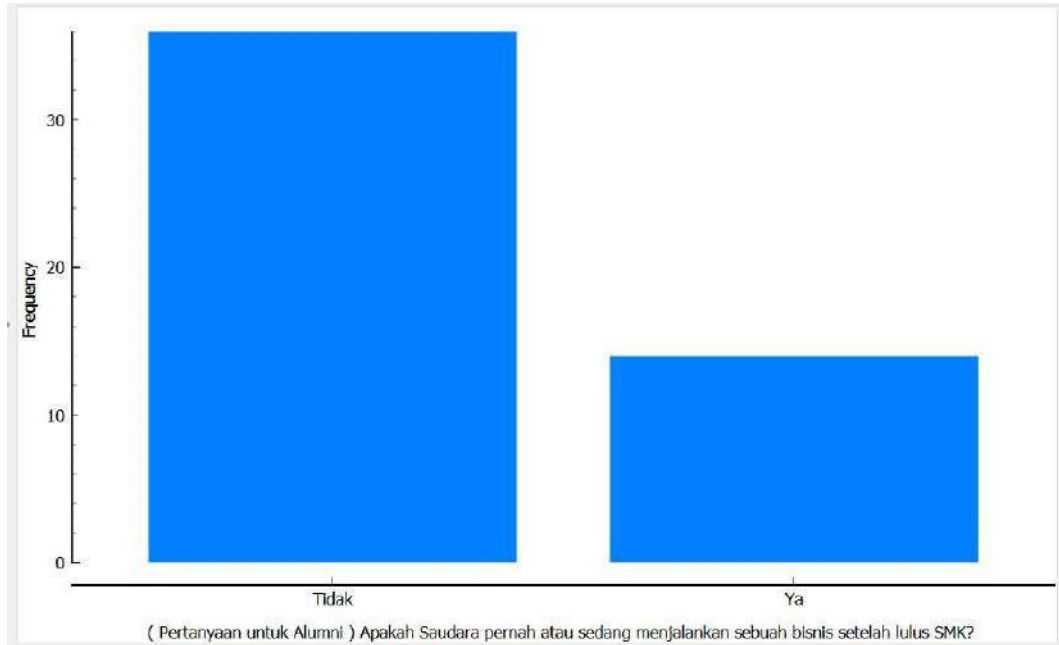


Gambar 3.6 Informasi Jumlah Kolom

Pada gambar 3.6, dengan menggunakan aplikasi *Orange Data Mining*, hasil kuisisioner dan data nilai akan dipanggil melalui *File Reader*, dan akan dilakukan pengecekan data melalui menu *preprocessing* untuk mengatasi data yang kosong.

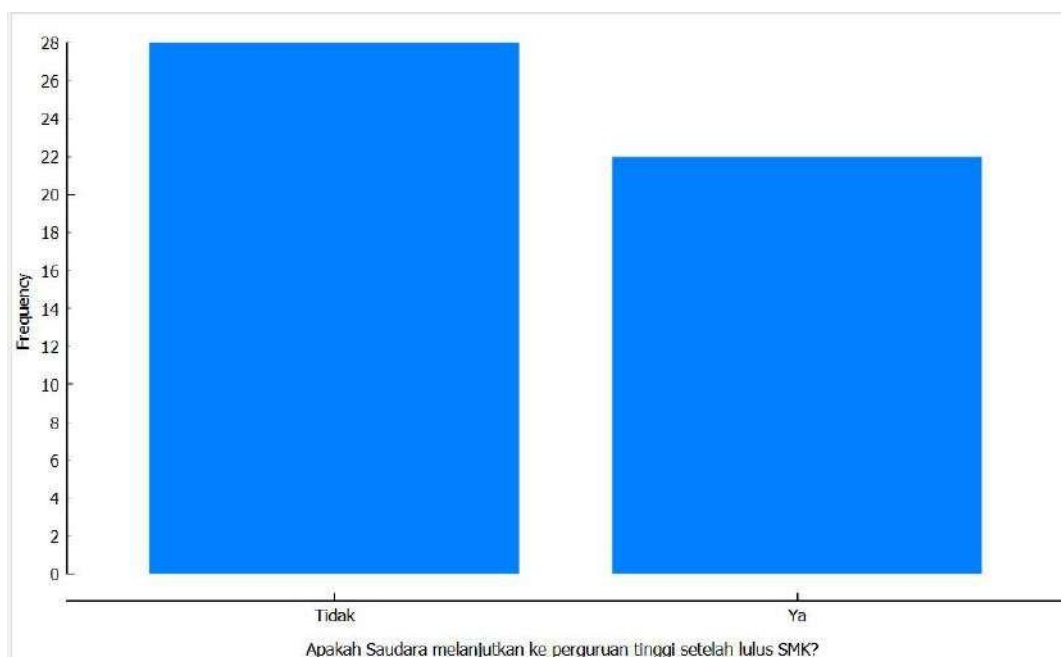
Untuk melakukan *EDA*, dapat digunakan melihat sebaran data melalui grafik *histogram*. Dengan menggunakan *widget Distribution* pada aplikasi *Orange Data Mining*, tergambar grafik histogram yang dapat diketahui sebaran data

alumni yang berwirausaha, melanjutkan pendidikan dan bekerja. Pada grafik sebaran data alumni yang berwirausaha, menunjukkan bahwa sebagian besar menyatakan tidak berwirausaha.



Gambar 3.7 Grafik Sebaran Berwirausaha

Pada gambar 3.7, grafik sebaran data alumni berwirausaha, menunjukkan bahwa sebagian besar menyatakan tidak berwirausaha.

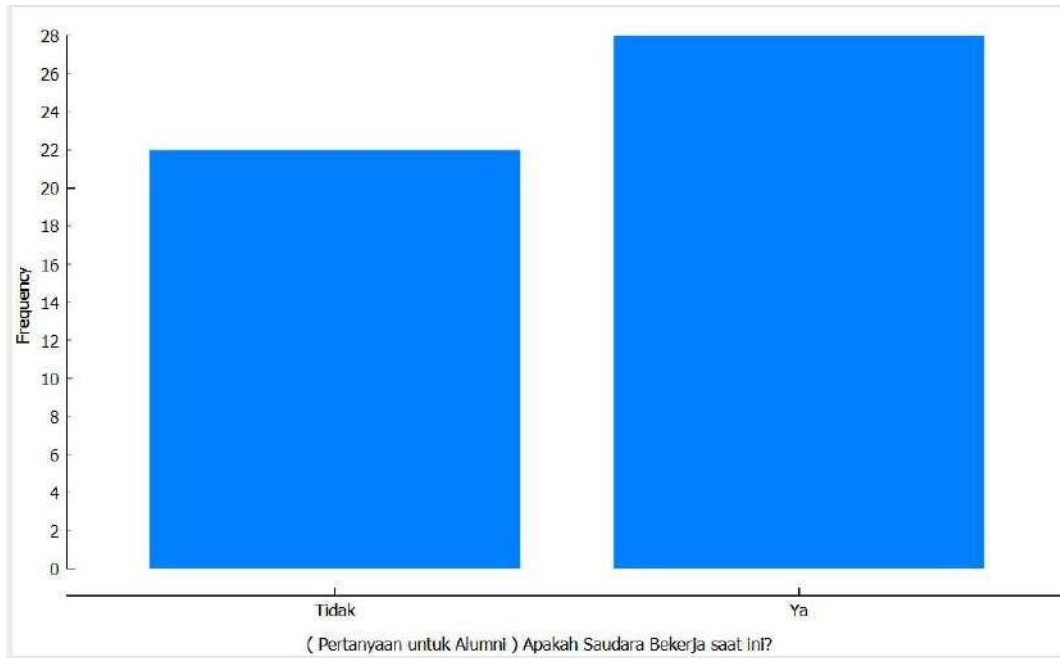


Gambar 3.8 Grafik Sebaran Alumni Kuliah

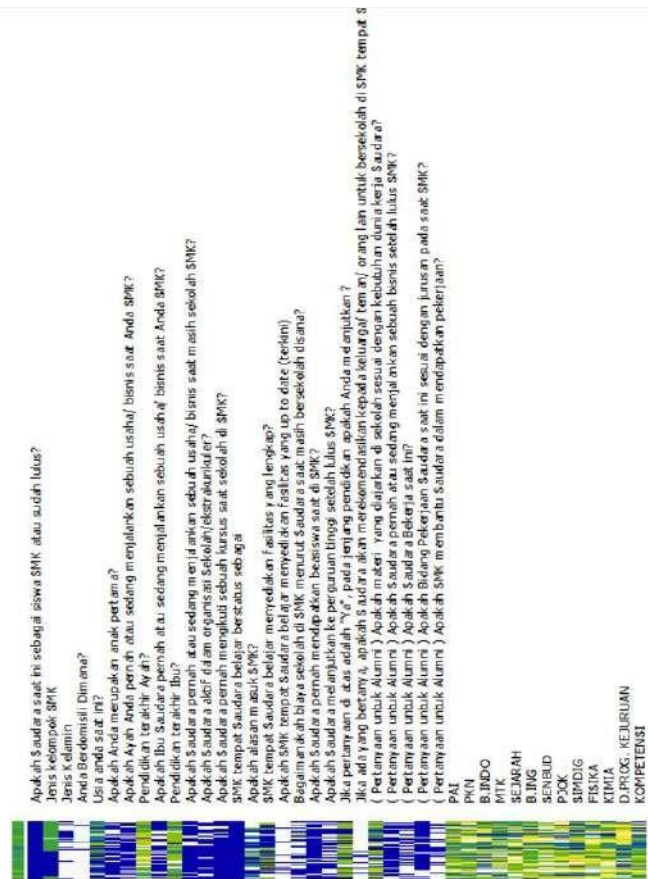
Pada gambar 3.8 sebaran data alumni yang melanjutkan pendidikan, frekuensi alumni yang berminat melanjutkan pendidikan lebih rendah dibanding yang tidak melanjutkan kuliah.

Pada gambar 3.9, frekuensi alumni yang bekerja saat ini lebih tinggi dibandingkan yang tidak bekerja. Selain melihat sebaran data, ada salah satu aspek penting dalam EDA yaitu memahami hubungan antara variabel - variabel dalam dataset, dan ini termasuk menganalisis korelasi.

Gambar 3.10 Grafik Alumni Bekerja



Gambar 3.9 Grafik Sebaran Alumni Bekerja



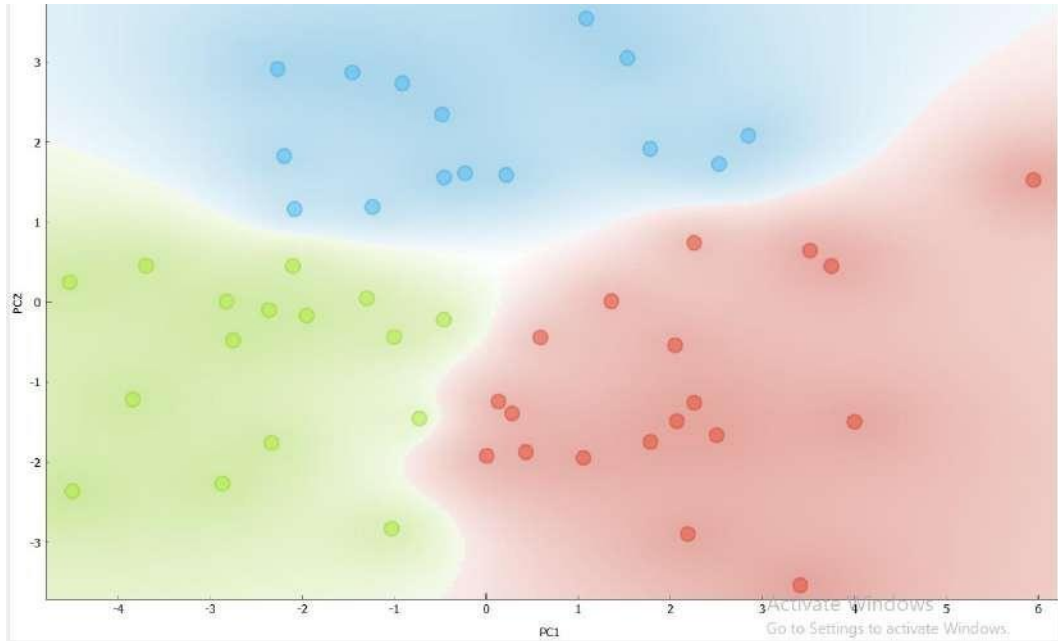
Gambar 3.10 Heatmap

Pada gambar 3.10 didapat hasil korelasi dengan Heatmap. Heatmap dengan skala warna, biru sering kali diasosiasikan dengan nilai yang lebih rendah atau dingin. Hijau dengan nilai-nilai yang sedang atau sedang di tengah rentang korelasinya. Sedangkan, dengan skala warna kuning, menggambarkan nilai dengan korelasi tinggi.

*Exploratory Data Analysis (EDA)* karena tujuan utamanya adalah untuk memahami dan mengidentifikasi fitur-fitur yang paling informatif atau berpengaruh dalam *dataset*. *Feature ranking* membantu mendapatkan wawasan tentang kontribusi setiap fitur terhadap tujuan analisis. *Feature ranking* dalam *machine learning* digunakan untuk mengidentifikasi dan mengurutkan fitur-fitur (variabel independen) yang paling relevan atau memiliki kontribusi paling signifikan Pemilihan Fitur Berbasis Domain: *Feature ranking* yang ditunjukkan dalam *scatter plot* dapat digunakan untuk memvalidasi fitur-fitur yang dipilih berdasarkan pengetahuan domain atau asumsi bisnis. Jika fitur yang dianggap

penting oleh ahli domain tidak mendapatkan peringkat yang tinggi, mungkin ada kesalahan dalam pemahaman tentang data atau diperlukan analisis lebih lanjut.

Pada penelitian, alumni akan merekomendasikan SMK berdasarkan hasil kuisisioner dengan mempertimbangkan beberapa faktor utama, peringkatnya yaitu :



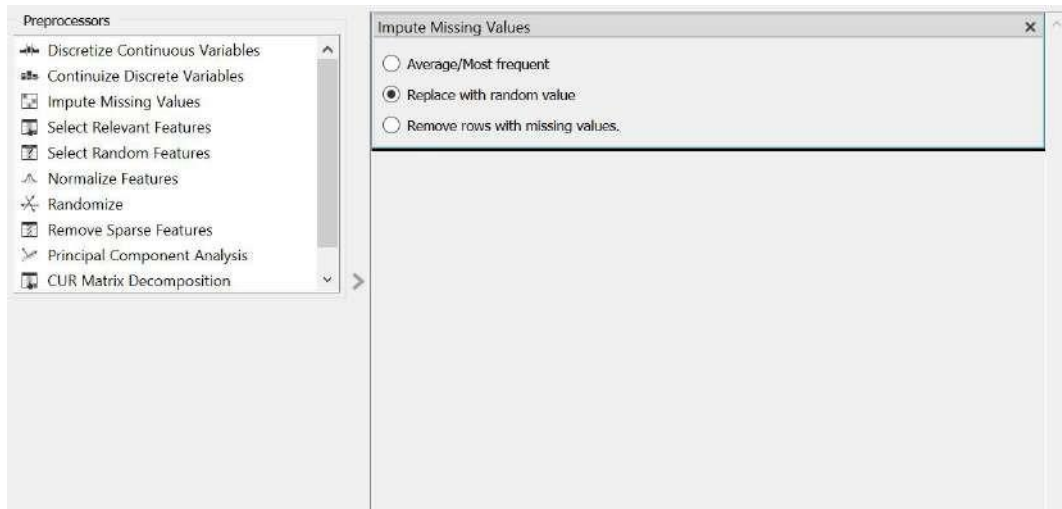
Gambar 3.11 Hasil Scatter Plot

Pada gambar 3.11 terdapat hasil *scatter plot* yang menunjukkan korelasi atau metode lainnya, sehingga dapat memberikan gambaran tentang fitur mana yang memiliki hubungan yang lebih kuat atau lebih lemah dengan variabel lainnya. Dengan menggunakan *scatter plot* dan *feature ranking* secara bersamaan, analisis data dapat mendapatkan pemahaman yang lebih lengkap tentang struktur dan karakteristik data.

### 3.2.4 PREPROCESSING

Pada tahap ini, dilakukan pemeriksaan pada data seperti mengganti data kosong dengan data minimal, maksimal ataupun rata – rata, menghapus data yang sama serta mengubah data kategori menjadi numerik. Proses machine learning dapat dilakukan jika atribut bernilai integer, sehingga atribut yang bernilai kategori harus dilakukan adanya perubahan.

### 3.2.4.1 MISSING VALUES



Gambar 3.12 Cek Missing Values

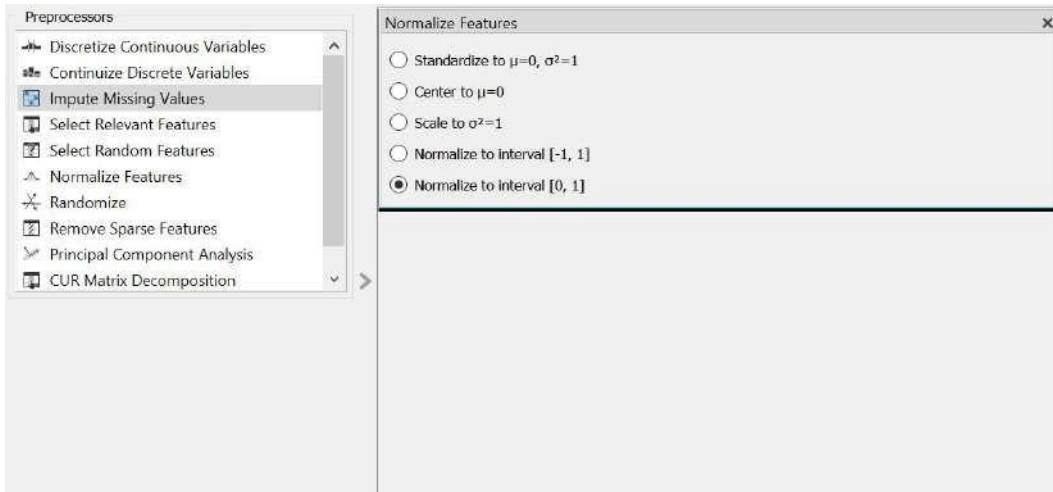
Pada gambar 3.12, dilakukan proses preprocessing dengan mengecek *missing values*. Sebelumnya, didapatkan bahwa nilai *missing values* 0.1%.

*Missing values* ini disebabkan oleh inputan dari alumni yang tidak lengkap dan *filter form* yang tidak dibuat, sehingga peserta didik alumni dapat melewati isian tanpa harus mengisi isian terlebih dahulu. *Missing values* dapat diatasi dengan *Replace with random values*.

Imputasi *missing values* dengan menggantikan *replace random values* pada Orange adalah metode dengan nilai yang hilang dalam dataset diisi nilai acak dari distribusi yang sesuai. Dengan menggunakan metode ini, dipertahankan sejumlah sifat statistik dari data yang ada, dan mengatasi ketidaklengkapan nilai - nilai yang hilang.



### 3.2.4.2 NORMALISASI

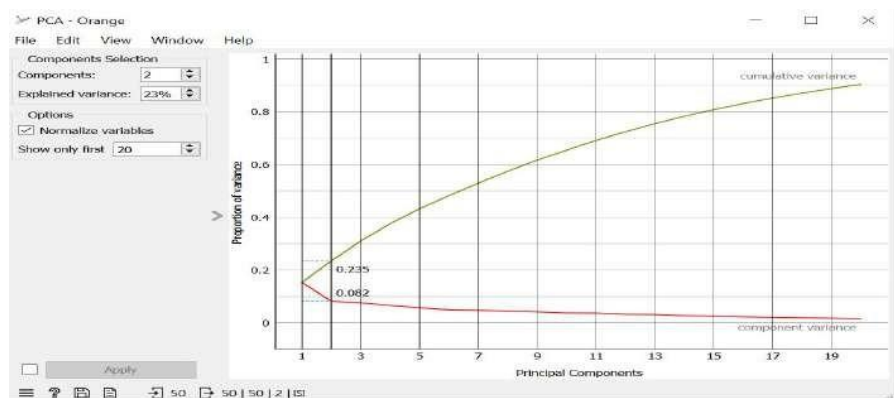


Gambar 3.13 Normalisasi

Pada gambar 3.13, dilakukan proses normalisasi. Proses ini mengubah mengubah setiap nilai dalam suatu atribut sehingga berkisar antara 0 dan 1 setelah sebelumnya dilakukan pengubahan data *categorical* menjadi *numeric*.

### 3.2.5 PCA

PCA ( *Principal Component Analysis* ) digunakan untuk mereduksi data menjadi 2 komponen yaitu hasil kuisioner dari beberapa pertanyaan kepada alumni dan nilai peserta didik dari semua pelajaran. Principal Component Analysis merupakan teknik yang biasa digunakan untuk menyederhanakan suatu data, dengan cara mentransformasi data secara linier sehingga terbentuk sistem koordinat baru dengan varians maksimum.



### Gambar 3.14 Penentuan Komponen PC

Pada gambar 3.14, Jumlah komponen yang sesuai adalah titik di mana pertambahan varian yang dijelaskan mulai melambat, membentuk "siku" pada grafik, dan yang paling mendekati 2. Grafik PCA dengan 2 komponen menunjukkan representasi data dalam ruang 2 dimensi setelah reduksi dimensi PCA. Grafik ini dapat membantu dalam memahami seberapa baik dua komponen utama dapat menjelaskan data yaitu Data Kuisisioner dan Data Nilai.

Ketika *explained variance* sebesar 23%, ini berarti dua komponen utama yang dipertimbangkan dalam PCA dapat menjelaskan 23% dari variasi total dalam data. Principal komponen kedua (PC2) mungkin memberikan kontribusi sebagian dari nilai ini. Nilai 23% ini ditentukan otomatis oleh sistem.

*Proporsi variance* antara 0,082 dan 0.235 yang mendekati 1 menunjukkan seberapa besar variasi yang dijelaskan oleh 2 elemen utama. Nilai yang lebih tinggi menunjukkan kemampuan 2 elemen tersebut untuk menjelaskan lebih banyak variasi dalam data, tetapi nilai ini hanya merupakan persentase kecil dari total variasi, yang berarti bahwa lebih banyak elemen diperlukan untuk menjelaskan sebagian besar variasi, artinya jumlah data alumni masih sedikit.

### 3.2.6 CLUSTERING

Untuk menentukan kluster dengan menggunakan perhitungan manual dan silhouette score, berikut adalah langkah-langkah yang dapat diikuti:

1. Inisialisasi Kluster:
  - a. Pilih jumlah kluster yang ingin dibentuk (misalnya,  $k=2$ ,  $k=3$ , dst.).
  - b. Inisialisasi pusat kluster secara acak.
2. Hitung Jarak:

Hitung jarak antara setiap data point dan pusat kluster. Anda dapat menggunakan metrik jarak seperti Euclidean distance  $d_{ij} = \sqrt{(x_i - c_{jx})^2 + (y_i - c_{jy})^2}$
3. Menandai Kluster:

Alokasikan setiap data point ke kluster dengan pusat terdekat.

**4. Menghitung Silhouette Score:**

a. Untuk setiap data point dalam kluster, hitung silhouette score menggunakan rumus:  $s(i) = \max\{a(i), b(i)\} - a(i)$

$s(i)$ : Silhouette score untuk data point ke-i.

$a(i)$ : Rata-rata jarak antara data point ke-i dan semua data point dalam kluster yang sama.

$b(i)$ : Rata-rata jarak antara data point ke-i dan semua data point dalam kluster lainnya (yang memiliki jarak minimum dengan data point ke-i).

**5. Menghitung Silhouette Score Total:**

Hitung rata-rata dari semua silhouette score untuk mendapatkan silhouette score total kluster.

**6. Iterasi:**

Ulangi langkah 2-5 untuk setiap iterasi hingga konvergensi (ketika tidak ada perubahan dalam alokasi kluster).

**7. Pilih Kluster dengan Silhouette Score Tertinggi:**

Jumlah Kluster	Silhouette Score
2	0.5
3	0.6
4	0.55
5	0.52

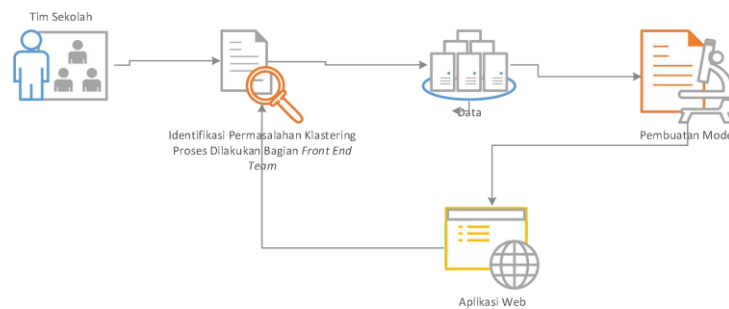
Tabel 3.2 Penentuan Kluster

Pada tabel 3.2 jumlah kluster didapat dari Silhouette Score tertinggi yaitu kluster 3.

### 3.2.7 PERANCANGAN *DEPLOY WEB*

#### 3.2.7.1 ARSITEKTUR SISTEM

Arsitektur sistem dalam *machine learning* merujuk pada struktur dan organisasi keseluruhan dari sistem yang digunakan untuk mengembangkan, melatih, dan menerapkan model machine learning. Arsitektur sistem ini mencakup beberapa komponen dan tahapan yang diperlukan dalam siklus hidup model machine learning.

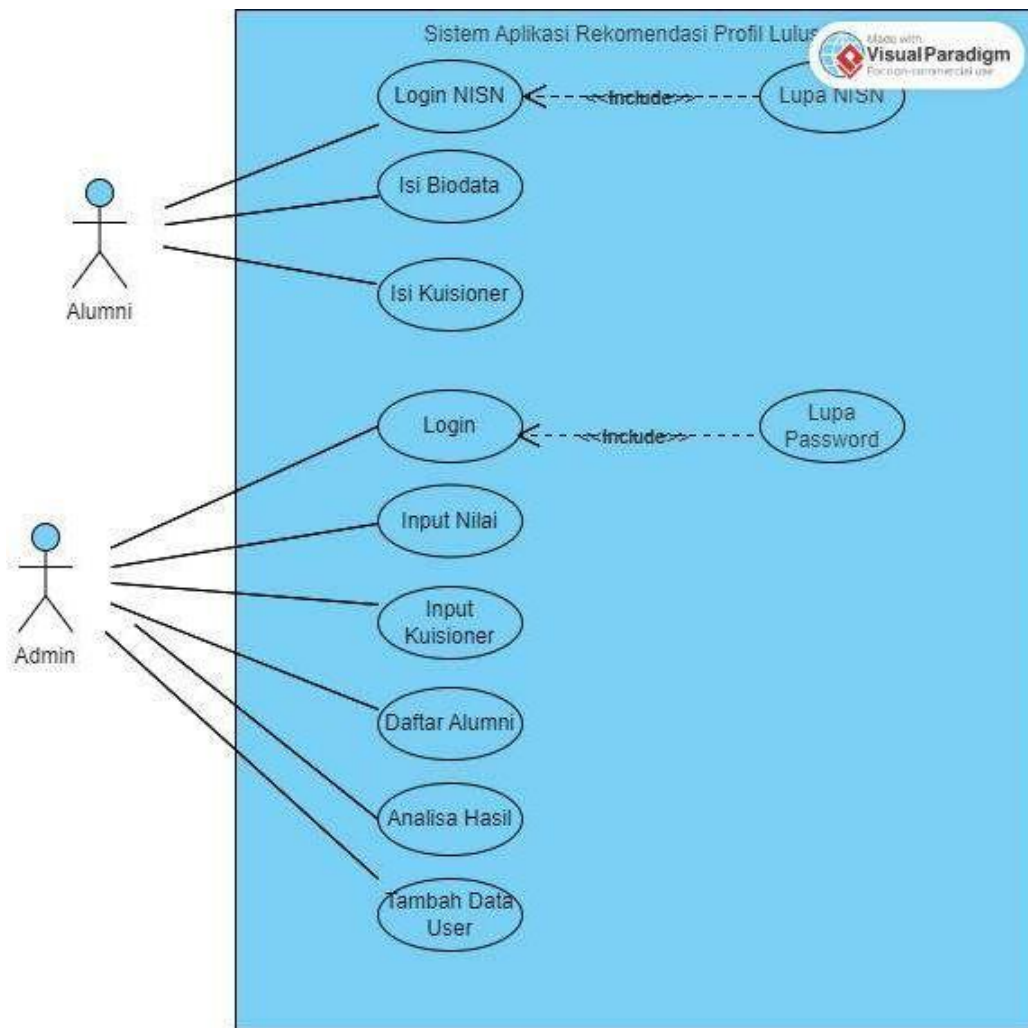


Gambar 3.15 Arsitektur Sistem

Pada gambar 3.15, arsitektur sistem dikelola oleh Tim Sekolah melalui identifikasi permasalahan melalui klustering dengan melalui pengisian kuisisioner. Selanjutnya, hasil akan disimpan dalam bentuk *data*. Tahapan berikutnya, adalah pembuatan model yang akan ditampilkan dalam aplikasi web. *Dashboard* akan dirancang dengan *Python*. Perancangan design dashboard terdiri dari dua halaman yaitu halaman beranda / index, dan halaman hasil proses K - Means. Halaman beranda berisi informasi akun admin dan input file dataset dengan format csv.

#### 3.2.7.2 *USE CASE DIAGRAM*

Use case diagram diagram digunakan untuk membantu dalam merancang arsitektur sistem dengan menyoroti hubungan dan interaksi antara berbagai bagian dari sistem. Hal ini memungkinkan pengembang untuk merencanakan struktur dan organisasi sistem secara lebih baik.

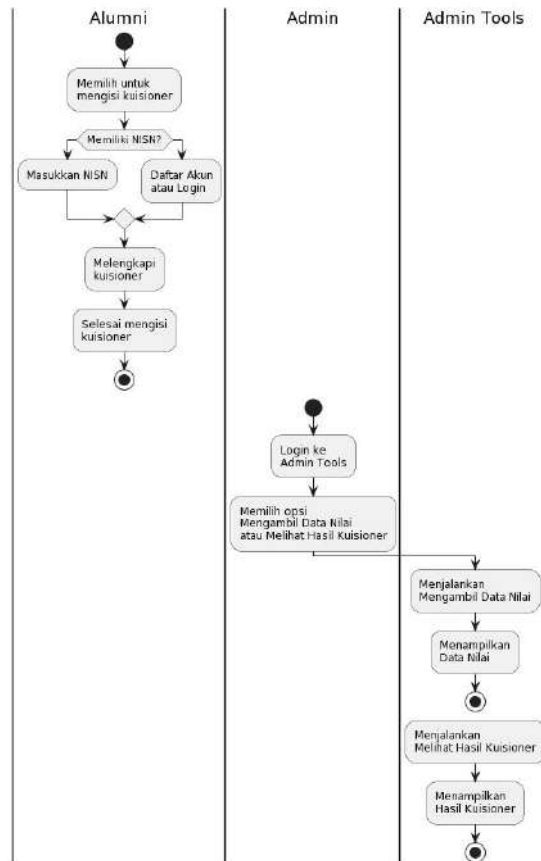


Gambar 3.16 Use Case Diagram

Sesuai dengan diagram *Use Case* pada gambar 3.16 , dapat dibuat *Dashboard* dengan 2 *login* untuk aktor yaitu admin ataupun alumni. Alumni dapat memasukkan informasi identitas dasar mereka ke dalam sistem berupa NISN ( Nomor Induk Siswa Nasional ). Alumni dapat menginput kuisisioner dan nilai melalui formulir yang disediakan. Admin juga dapat memasukkan kuisisioner dan nilai ke dalam sistem. Admin dapat mengelola kuisisioner yang dapat diberikan kepada alumni, dan juga admin dapat mendaftarkan alumni baru ke dalam sistem.

### 3.2.7.3 ACTIVITY DIAGRAM

Diagram aktivitas membantu dalam analisis dan perancangan sistem dengan memodelkan aktivitas dan alur kerja yang terlibat dalam suatu sistem atau komponen sistem.

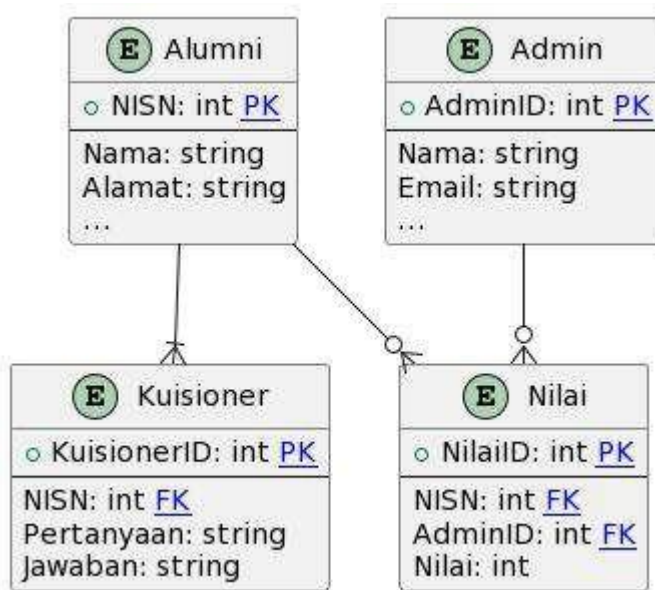


Gambar 3.17 Activity Diagram

Pada gambar 3.17 *activity diagram*, alumni memulai dengan menginput biodata dan mengisi melalui kuisisioner, selanjutnya admin mengelola kuisisioner dan menganalisis data alumni dan menerapkan metode *machine learning*.

### 3.2.7.4 ENTITY RELATIONSHIP DIAGRAM ( ERD )

ERD membantu dalam perencanaan dan pengembangan sistem dengan memberikan gambaran yang jelas tentang bagaimana data diorganisir dan dikelola dalam suatu sistem.



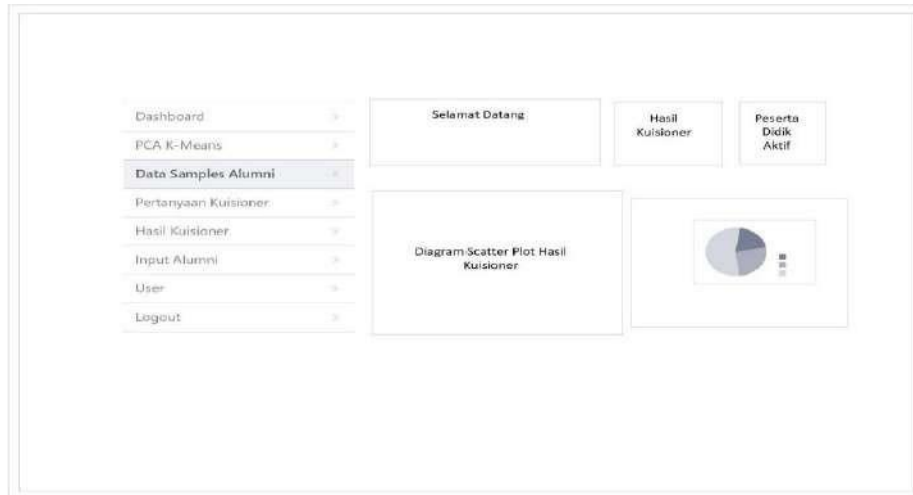
Gambar 3.18 ERD

Pada gambar 3.18, ERD ( *Entity-Relationship Diagram* ) yang terdiri dari empat tabel yaitu "Alumni", "Admin", "Kuisiner", dan "Nilai" dengan kunci utama (primary key) "NISN" (Nomor Induk Siswa Nasional) untuk tabel "Alumni", "AdminID" untuk tabel "Admin", "KuisinerID" untuk tabel "Kuisiner", dan "NilaiID" untuk tabel "Nilai" akan menggambarkan hubungan antara entitas-entitas tersebut.

### 3.2.7.5 PERANCANGAN INTERFACE

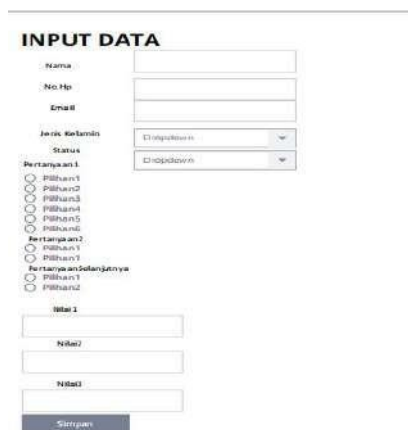
Gambar 3.19 Perancangan Halaman Depa

Pada gambar 3.19, ada perancangan halaman login yang terdiri dari *username* alumni dan admin. Langkah pertama, *username* alumni harus mendaftar terlebih dahulu melalui link Daftar



Gambar 3.20 Pemilihan Menu Pada Sistem

Pada gambar 3.20, perancangan dashboard Sistem Rekomendasi Profil Lulusan SMK Negeri 6 Pekanbaru yang terdiri dari beberapa menu yaitu PCA K – Means, Data Sample / Alumni, Rekomendasi, Pertanyaan Kuisisioner, Hasil Kuisisioner, Input Alumni dan User bagi menu *Admin*. Sedangkan bagi user alumni, menu yang ada adalah Daftar User dan Input Kuisisioner.



Gambar 3.21 Input Data Alumni



Pada gambar 3.21, baik user alumni ataupun alumni dapat memasukkan data kuisisioner rekomendasi profil SMK dan nilai.

Alumni / Daftar Alumni					
No	Nama	Jenis Kelamin	Status	Kontak	Action
1					<a href="#">Lihat Data</a>
2					<a href="#">Lihat Data</a>
3					<a href="#">Lihat Data</a>

Gambar 3.22 Daftar Sample / Alumni

Pada gambar 3.22, Admin dapat melihat data alumni yang sudah mengisi data kuisisioner.

## **BAB IV**

### **EKSPERIMEN DAN ANALISIS**

#### **4.1 PARAMETER EKSPERIMEN**

Parameter yang akan digunakan pada penelitian ini, akan menggunakan metode K – Means dengan jumlah kluster sebanyak 3 dengan penginisialisasi acak, dengan pengukuran jarak *Euclidean Distance*.

#### **4.2 KARAKTERISTIK DATA**

Dengan metode K – Means, jumlah data sebanyak 50 *record* , dengan 38 atribut, tidak ada *missing values*, terdapat korelasi antar atribut satu dengan yang lain.

#### **4.3 SPESIFIKASI PERALATAN UJI COBA**

Penelitian ini menggunakan Orange Data Mining dan pembuatan Dashboard dengan menggunakan CodeIgniter dengan API dari Tableau. Perangkat fisik yang diperlukan adalah Sistem Operasi: Windows 7/8/10 (64-bit) atau macOS dengan processor Intel Core i5 atau setara (disarankan Intel Core i7 atau setara dengan RAM Minimal 8 GB (disarankan 16 GB atau lebih), Ruang Penyimpanan minimal 10 GB ruang bebas pada hard drive, Resolusi Layar Minimal 1366x768 piksel, koneksi internet yang lancar.

#### **4.4 TAHAPAN MACHINE LEARNING PADA PYTHON**

Tahapan machine learning dimulai dari proses pengumpulan data, persiapan data, memilih pemodelan, uji coba pemodelan, evaluasi pemodelan, peningkatan pemodelan dan menentukan prediksi.

#### 4.4.1 DATA ACQUISITION

Dataset dalam penelitian ini menggunakan dataset dari hasil rekapan nilai dari

dari alumni yang menjadi peserta didik pada tahun sebelumnya dan hasil kuisisioner yang menjadi Dataset.

#### 4.4.2 DATA EXPLORATION

Proses data exploration merupakan proses mengeksplorasi dataset, yaitu melihat isi dataset, apakah mempunyai *missing values*, kemudian melakukan transformasi data dari kategori menjadi numerik.

##### a. Menyiapkan Library dan Data

```
[ ] # import tools
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.cluster import KMeans
```

Gambar 4.1 Import Library di Python

Pada gambar 4.1, Proses analisis dengan menggunakan *Python* dimulai dengan menyiapkan *library* yang akan dipakai. Library yang digunakan adalah *NumPy*, *pandas*, *matplotlib*, *seaborn*, *sklearn*. Library *Numpy* digunakan untuk data *analysis tools*, library *matplotlib* dan *seaborn* untuk visualisasi data serta *library scikit – learn* untuk *machine learning*.

```
[ ] from google.colab import files
uploaded = files.upload()

Choose Files No file chosen Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.
Saving tesisver1.csv to tesisver1.csv

[ ] # import data
df = pd.read_csv('tesisver1.csv')
df.head()
```

Gambar 4.2 Memanggil Dari Dataset

Pada gambar 4.2, contoh penggunaan dari dua fungsi di atas *upload* dan *read\_csv*. Pertama, fungsi *upload\_data()* digunakan untuk membuat *DataFrame* dari data contoh. Kemudian, fungsi *read\_csv\_file()* digunakan untuk membaca *file CSV* dari path yang diberikan ke dalam *DataFrame*.

#### 4.4.3 DATA PRE- PROCESSING

Pada proses pre – processing dilakukan beberapa tahapan, yaitu :

##### a. Data Cleaning

Proses *data cleaning* merupakan proses pembersihan data dari nilai – nilai yang kosong atau missing value, pada proses ini juga akan dilakukan proses pengecekan data yang tidak lengkap atau tidak relevan, untuk selanjutnya dilakukan proses perbaikan data.

##### 1. Missing Value

Missing value dapat dilakukan dengan :

##### Drop Missing Value



```
df.isnull().sum()
Label
Jurusan Anda di SMK 6 Pekanbaru      0
Jenis Kelamin                        0
Anda Berdomisili Dimana?             0
Usia anda saat ini?                  0
Apakah Anda merupakan anak pertama? 0
Apakah Ayah Anda pernah atau sedang menjalankan sebuah usaha/ bisnis saat Anda SMK? 0
Pendidikan terakhir Ayah?           0
Apakah Ibu Saudara pernah atau sedang menjalankan sebuah usaha/ bisnis saat Anda SMK? 0
Pendidikan terakhir Ibu?             0
Apakah Saudara pernah atau sedang menjalankan sebuah usaha/ bisnis saat masih sekolah SMK? 0
Apakah Saudara aktif dalam organisasi Sekolah/ekstrakurikuler?                 0
Apakah Saudara pernah mengikuti sebuah kursus saat sekolah di SMK?          0
Apakah alasan masuk SMK?              0
SMK tempat Saudara belajar menyediakan fasilitas yang lengkap?               0
Apakah SMK tempat Saudara belajar menyediakan fasilitas yang up to date (terkini)? 0
Bagaimanakah biaya sekolah di SMK menurut Saudara saat masih bersekolah disana? 0
Apakah Saudara pernah mendapatkan beasiswa saat di SMK?                      0
Apakah Saudara melanjutkan ke perguruan tinggi setelah lulus SMK?             0
Jika pertanyaan di atas adalah "Ya", pada jenjang pendidikan apakah Anda melanjutkan ? 0
( Pertanyaan untuk Alumni ) Apakah Saudara pernah atau sedang menjalankan sebuah bisnis setelah lulus SMK? 0
( Pertanyaan untuk Alumni ) Apakah Saudara Bekerja saat ini?                   0
( Pertanyaan untuk Alumni ) Apakah Bidang Pekerjaan Saudara saat ini sesuai dengan jurusan pada saat SMK? 0
( Pertanyaan untuk Alumni ) Apakah SMK membantu Saudara dalam mendapatkan pekerjaan? 0
```

Gambar 4.3 Cek Missing Value

Pada gambar 4.3, dari proses cek *missing value*, ditemukan nilai *missing values* 0.

2. Setelah dilakukan pengecekan missing value dan ditemukan baris yang mengandung missing value, maka dilakukan proses penghapusan baris yang mengandung missing value dengan perintah sebagai berikut:

##### 3. Hapus Kolom

Penghapusan atribut yang tidak digunakan dalam proses *machine learning*, dilakukan karena atribut yang tidak digunakan karena datanya bernilai primary.

##### 4. Cek Data Ganda

Dari hasil pengecekan data ganda, maka ditemukan terdapat baris data ganda, sehingga data ganda harus dilakukan penghapusan.

```
[10] df.duplicated().value_counts()

False    44
True      3
dtype: int64
```

Gambar 4.4 Jumlah Data Ganda

Pada gambar 4.4, dari hasil pengecekan data ganda, maka ditemukan data ganda sebesar 3, selanjutnya dilakukan proses penghapusan data ganda, sehingga dataset yang digunakan untuk proses machine learning sejumlah 44 baris.

#### b. Exploratory Data Analysis ( EDA )

Salah satu bentuk dalam melakukan analisis data adalah dengan melakukan *exploratory data analysis* ( EDA ). Pada EDA, dilakukan eksplorasi data sehingga akan mendapatkan informasi yang tepat terkait dataset. EDA merupakan proses untuk mempelajari bentuk data yang akan menentukan bagaimana proses pengolahan datanya tersebut. Pada tahap ini, dilakukan pemeriksaan pada data seperti mengganti data kosong dengan data minimal, maksimal ataupun rata – rata, menghapus data kategori menjadi numerik. Hasil pengecekan informasi dari dataset yang akan digunakan seperti tampilan berikut

#### 1. Memeriksa Kolom Numerik

Gambar 4.5 Cek Data Type

Pada gambar 4.5, dari hasil penelitian, ditemukan semua atribut sudah bertipe data float ataupun integer.

## 2. Memeriksa Jumlah Data Points

```
[8] df.shape
(47, 39)
```

Gambar 4.6 Cek Jumlah Data Points

Pada gambar 4.6, jumlah kolom dalam *DataFrame* ( `df.shape` ) mengindikasikan kebutuhan fitur atau atribut yang ada dalam dataset sehingga dapat diketahui kompleksitas dataset dan menganalisis hubungan antara fitur-fitur tersebut.

```
[9] df.describe()
```

	Jurusan Anda di SMK 6 Pekanbaru	Jenis Kelamin	Anda Berdomisili Di mana?	Usia anda saat ini?	Apakah Anda merupakan anak pertama?	Apakah Ayah Anda pernah atau sedang menjalankan sebuah usaha/ bisnis saat Anda SMK?	Pendidikan terakhir Ayah?	Apakah Ibu Saudara pernah atau sedang menjalankan sebuah usaha/ bisnis saat Anda SMK?	Pendidikan terakhir Ibu?	Apakah Saudara pernah atau sedang menjalankan sebuah usaha/ bisnis saat masih sekolah SMK?	...	MTK	SEJARAH	B.ING	SENBA
count	47.000000	47.000000	47.000000	47.000000	47.000000	47.000000	47.000000	47.000000	47.000000	...	47.000000	47.000000	47.000000	47.000000	
mean	0.659574	0.723404	0.042553	0.304965	0.659574	0.680851	0.250000	0.574468	0.164894	0.829787	...	0.591017	0.668335	0.656383	0.7212
std	0.362755	0.452151	0.204030	0.301604	0.478975	0.471186	0.303959	0.499769	0.267215	0.379883	...	0.300167	0.280617	0.244864	0.2744
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.0000
25%	0.500000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	...	0.333333	0.529412	0.500000	0.5000
50%	0.500000	1.000000	0.000000	0.333333	1.000000	1.000000	0.250000	1.000000	0.000000	1.000000	...	0.611111	0.705882	0.700000	0.8000
75%	1.000000	1.000000	0.000000	0.333333	1.000000	1.000000	0.250000	1.000000	0.250000	1.000000	...	0.888889	0.882353	0.850000	0.9500
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	...	1.000000	1.000000	1.000000	1.0000

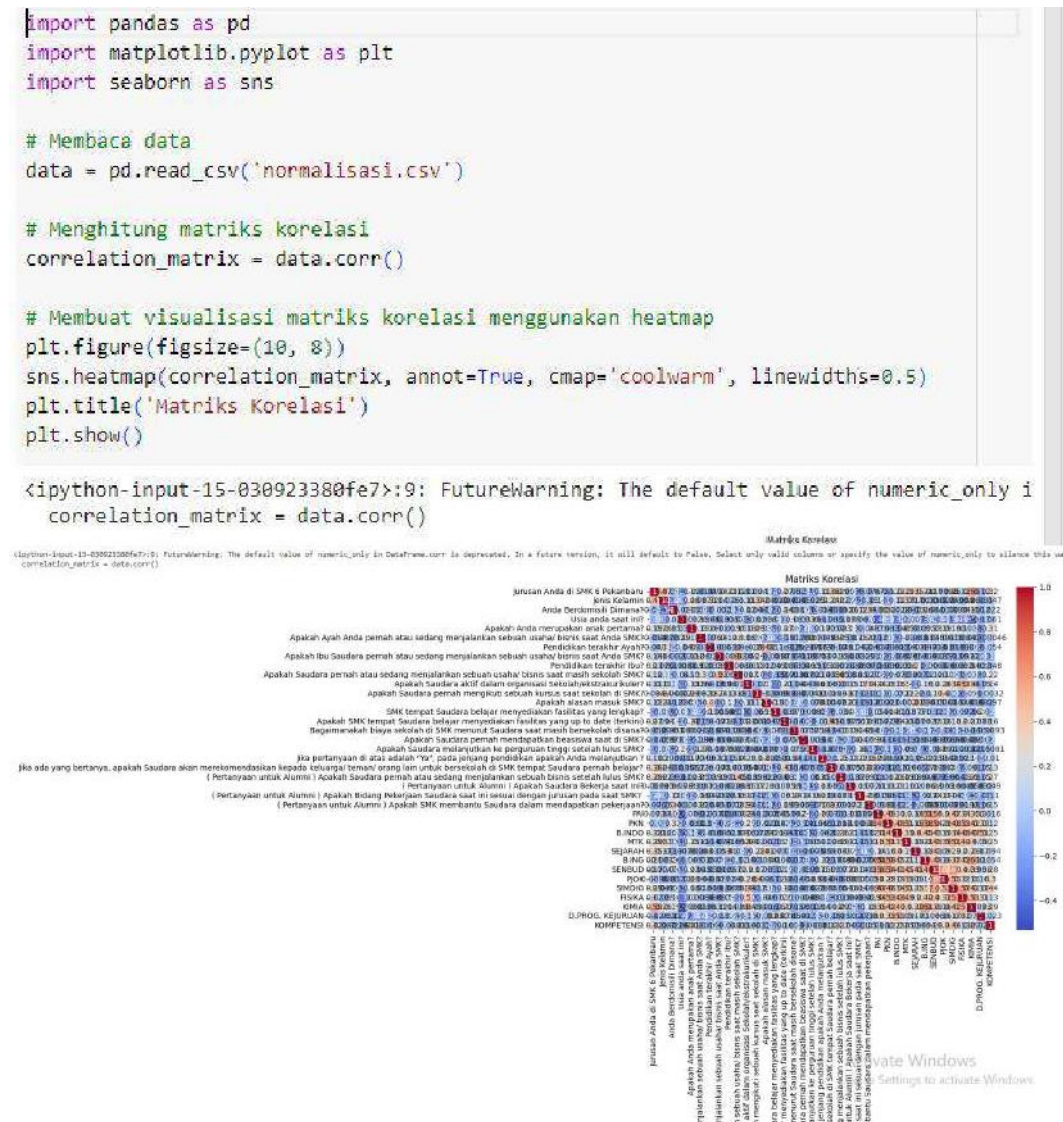
Gambar 4.7 Deskripsi Dataset

Pada gambar 4.7, `df.describe()` adalah metode pemrograman yang digunakan pada objek *DataFrame* dari *library* Pandas, biasanya digunakan dalam analisis data dengan Python. Fungsi `df.describe()` berguna untuk memberikan statistik deskriptif tentang data dalam *DataFrame*, seperti *mean*, deviasi standar, nilai minimum, kuartil, dan nilai maksimum setiap kolom data.

## 4. Melihat Informasi Pola Hubungan Antar Atribut

Hubungan korelasi dapat dilihat bahwa , antara atribut kuisioner dan nilai memiliki hubungan yang dekat, karena sama – sama bernilai mendekati 1, jadi nilai korelasi yang mendekati 1, maka hubungan antar atribut dekat, namun jika

nilai korelasi mendekati 0 bahkan ke minus ( - ) maka hubungan korelasi antar atribut semakin jauh.



Gambar 4.8 Korelasi

Pada gambar 4.8, Script ini menggunakan seaborn untuk membuat heatmap dari matriks korelasi. Argumen seperti `annot=True` digunakan untuk menampilkan nilai korelasi di dalam sel, `cmap='coolwarm'` mengatur skema warna, dan `linewidths=0.5` memberikan ketebalan garis antar sel. Akhirnya, `plt.show()` digunakan untuk menampilkan visualisasi.

#### 4.4.4 MODELLING DENGAN PYTHON DENGAN FLASK

Sebelum pemodelan, diperlukan integrasi antara Python, Database MySQL, dan Flask untuk menampilkan aplikasi berbasis web. Pada tahap modelling, penerapan algoritma untuk menentukan metode yang tepat.

```
from flask import Flask

def create_app():
    app = Flask(__name__)
    app.secret_key = 'your_secret_key'

    # Fungsi untuk membuat koneksi ke MySQL
    def get_mysql_connection():
        import mysql.connector
        return mysql.connector.connect(
            host='localhost',
            user='root',
            password='',
            database='dbuimel'
        )

    app.config['get_mysql_connection'] = get_mysql_connection

    # Register blueprints
    from app.controllers import login, admin, alumni, pdaktif, kuisioner, metode
    app.register_blueprint(login.login_bp)
    app.register_blueprint(admin.admin_bp)
    app.register_blueprint(alumni.alumni_bp)
    app.register_blueprint(pdaktif.paktif_bp)
    app.register_blueprint(kuisioner.kuisioner_bp)
    return app
```

Gambar 4.9 Konfigurasi Flask

Pada gambar 4.9, untuk membuat aplikasi *Flask* utama dan melakukan konfigurasi awal sebelum aplikasi dijalankan, fungsi *create\_app()* biasanya digunakan. Aplikasi ini dapat terdiri dari berbagai bagian, juga dikenal sebagai blueprints, yang disusun sesuai dengan fungsi dan logika tertentu.

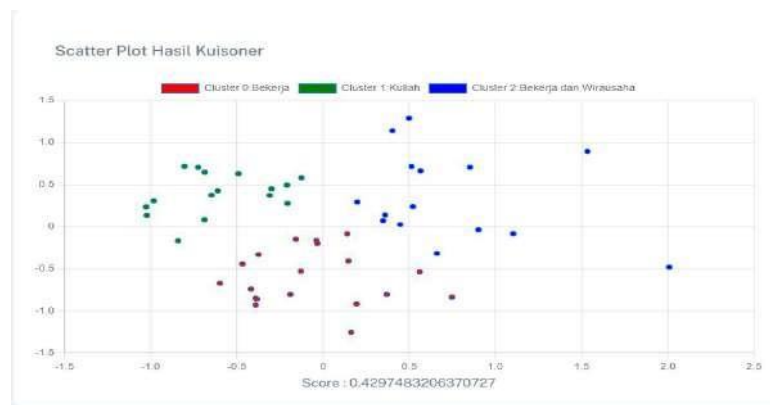
```
adminpy Client: 1 | adminpy Customized... | metodapy
> python3 app.py && python3 metodapy.py
1 from flask import Blueprint, render_template, session, current_app
2 import pandas as pd
3 from sklearn.cluster import KMeans
4 from sklearn.metrics import silhouette_samples
5 import numpy as np
6 from collections import Counter
7 from sklearn.decomposition import PCA
8 from sklearn.preprocessing import MinMaxScaler
9 from collections import defaultdict
10 import json
11 import copy
12 from sklearn.metrics import silhouette_score
13
14 def get_data_from_mysql():
15     # Buat koneksi ke database
16     conn = current_app.config['get_mysql_connection]()
17     cursor = conn.cursor(dictionary=True)
18     dataalumni_query = "SELECT dataalumni.status, nama as label, idalumni FROM dataalumni"
19     cursor.execute(dataalumni_query)
20     dataalumni = cursor.fetchall()
21
22     variabel_query = "SELECT variabel.isi, idvar FROM variabel WHERE variabel.idvar IN (1,2,4,5,6,7,8,9,10,11,23,24,25,26,27,28,29,30,31,32,33,34,35)"
23     cursor.execute(variabel_query)
24     variabel = cursor.fetchall()
25
26     data_from_database = []
27
28     for row in dataalumni:
29         dataheader = {}
30         dataheader['label'] = row['label']
31         dataheader['status'] = row['status']
32         for rowheader in variabel:
```

Gambar 4.10 Script K - Means dan PCA



Pada gambar 4.10, script tersebut merupakan bagian dari sebuah aplikasi yang menggunakan *Flask* yang memiliki banyak fungsi. Fungsi *get\_data\_from\_mysql* dapat digunakan untuk mengambil data dari database MySQL. Pertama, akan mengeksekusi permintaan untuk mendapatkan data alumni, serta status dan variabel yang diperlukan. Sebelum dikembalikan, data diproses dan disusun dengan benar

#### 4.5 EVALUASI MODELING K – MEANS DENGAN PYTHON



Gambar 4.11 Evaluasi Pengujian Python

Berdasarkan hasil penelitian, didapati bahwa jumlah klaster ada 3 dengan dominan bekerja, melanjutkan pendidikan, dan berwirausaha dengan hasil *score plot* 0.4297 dengan hasil rekomendasi

FASILITAS TERKINI	FASILITAS LENGKAP	KURIKULUM BUDI	BIDANG PEKERJAAN	INFO LOKER
Beberapa fasilitas mungkin masih kurang optimal, seperti keterbatasan ruang olahraga atau laboratorium tertentu. Kondisi ini mungkin memengaruhi pengalaman siswa dalam mendapatkan pengajaran praktis dalam beberapa mata pelajaran	Tidak Perlu ada peningkatan	Kurikulum di SMK Negeri 0 Pekanbaru didesain untuk mencakup aspek teori dan praktis. Adanya mata pelajaran yang relevan dengan dunia industri, penekanan pada keterampilan teknis, dan kerjasama dengan perusahaan memberikan siswa dasar yang kuat untuk memasuki dunia kerja.	Tidak Perlu ada peningkatan	Kebanyakan alumni tidak mendapatkan info lowongan pekerjaan dari sekolah hal ini perlu ditingkatkan
Beberapa fasilitas mungkin masih kurang optimal, seperti keterbatasan ruang olahraga atau laboratorium tertentu. Kondisi ini mungkin memengaruhi pengalaman siswa dalam mendapatkan pengajaran praktis dalam beberapa mata pelajaran	Tidak Perlu ada peningkatan	Tidak Perlu ada peningkatan	Tidak Perlu ada peningkatan	Kebanyakan alumni tidak mendapatkan info lowongan pekerjaan dari sekolah hal ini perlu ditingkatkan

Gambar 4.12 Hasil rekomendasi

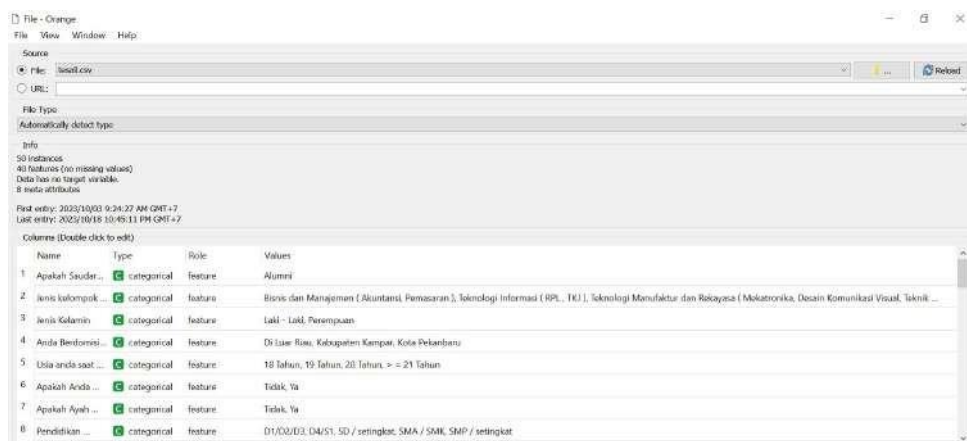
Pada gambar 4.12, didapat hasil dari kuisisioner yang dibagikan kepada alumni yang bekerja, melanjutkan pendidikan dan berwirausaha masih kurang dalam menyediakan fasilitas terkini, kurikulum yang disajikan oleh sekolah sama dengan dunia industri dan informasi lowongan masih kurang.

## 4.6 TAHAPAN MACHINE LEARNING DENGAN ORANGE

Untuk menguji validitas *script* Python, pada penelitian ini digunakan aplikasi Data Mining yaitu *Orange*.

### 4.6.1 DATA ACQUISITION DENGAN ORANGE

Untuk proses akusisi data, pada *Orange Application* dapat menggunakan Widget File. Widget File membaca file data input, yang dapat berupa tabel data yang memiliki instance data, dan kemudian mengirimkan dataset ke saluran outputnya. Widget ini juga memiliki direktori yang mengandung dataset sampel yang sudah pre-instal di Orange dan menyimpan catatan file yang terakhir dibuka. Widget dapat membaca data dari file Excel (.xlsx), file tab terbatas (.txt) sederhana, file comma-separated (.csv) atau URL. Informasi tentang format tambahan dapat ditemukan di bagian Format Lain di bawah ini.

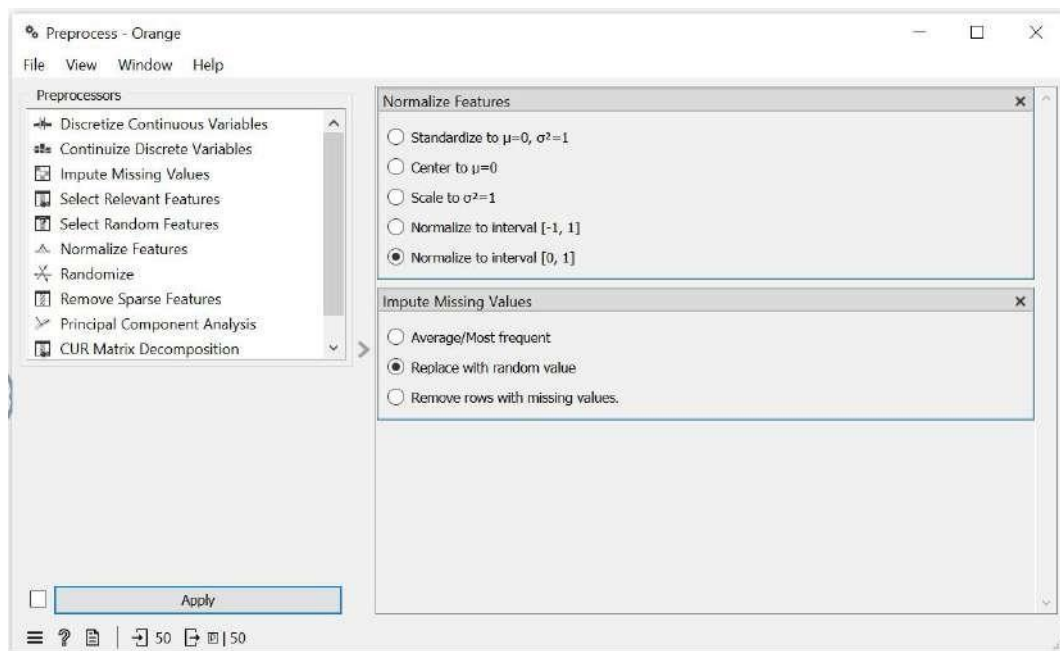


Gambar 4.13 *Widget File Orange*

Pada gambar 4.3, *Widget "File"* di Orange memungkinkan data dimuat dari file ke dalam alur kerja (*workflow*). *Widget* ini memungkinkan pengguna membaca data dari format file seperti CSV, Excel, dan SQL, serta menggunakan alat visual yang tersedia di Orange untuk menganalisis data.

#### 4.6.2 DATA EXPLORATION

*Data Exploration* adalah metode uji investigasi awal yang dimaksudkan untuk menemukan pola, anomali, hipotesis, dan asumsi. Dengan EDA pada Orange, pengguna dapat menemukan kesalahan awal, menemukan outlier, mengetahui hubungan antar data, dan menemukan faktor penting dalam data. Komponen EDA meliputi *preprocessing*, yang di dalamnya mencakup proses penanganan *missing values*, *outlier*, reduksi dimensi, transformasi, distribusi dan normalisasi data.

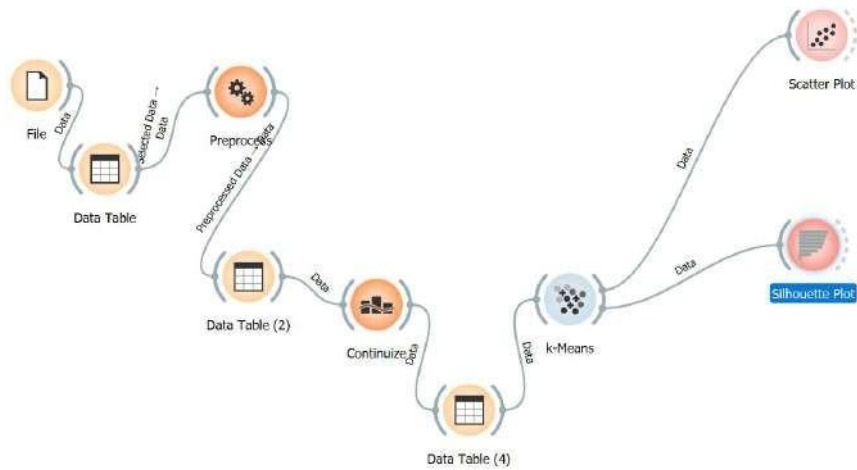


Gambar 4.14 *Data Exploration* pada Orange

Pada gambar, *normalize Widget* adalah *Widget* yang memungkinkan normalisasi data. Pilih kolom mana yang ingin dinormalisasi, dan *widget* akan mengubah nilai-nilai tersebut ke dalam rentang 0 hingga 1. Untuk mengisi atau

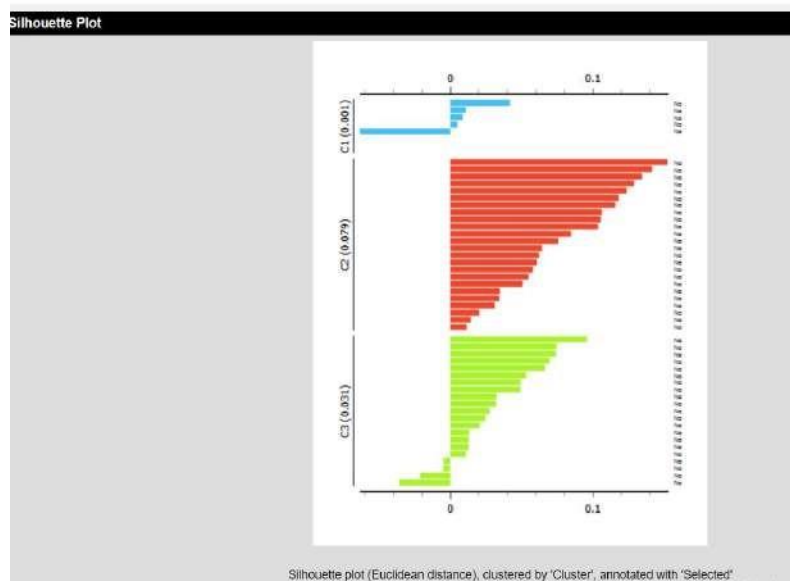
mengganti nilai yang hilang dalam kumpulan data disebut identitas nilai yang hilang dapat digunakan *widjet Impute* di *Orange*.

### 4.6.3 MODELLING PADA ORANGE



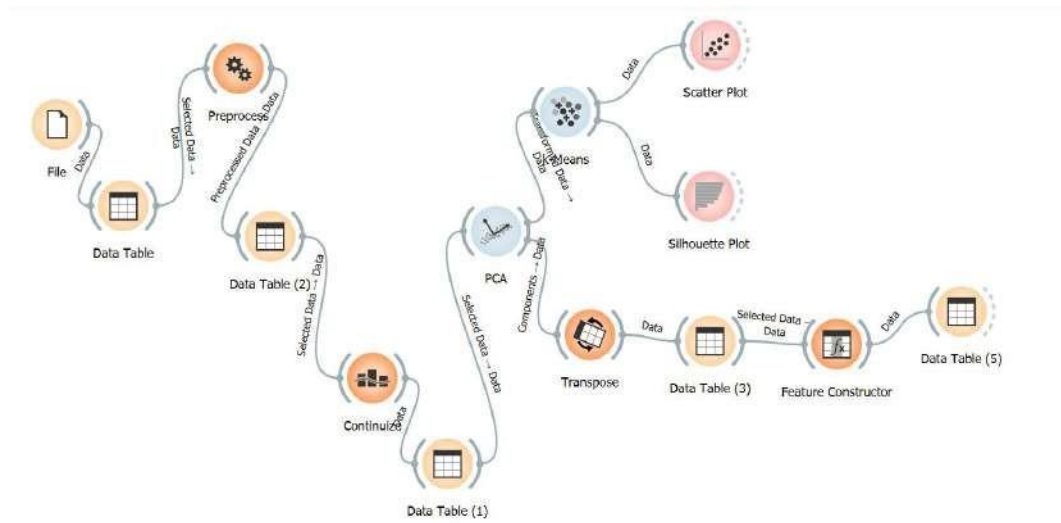
Gambar 4.15 K - Means dengan Orange

Pada gambar 4.15, hasil dari widget K-means Orange Data Mining biasanya memberikan informasi tentang pengelompokan atau kluster yang dihasilkan dari analisis.



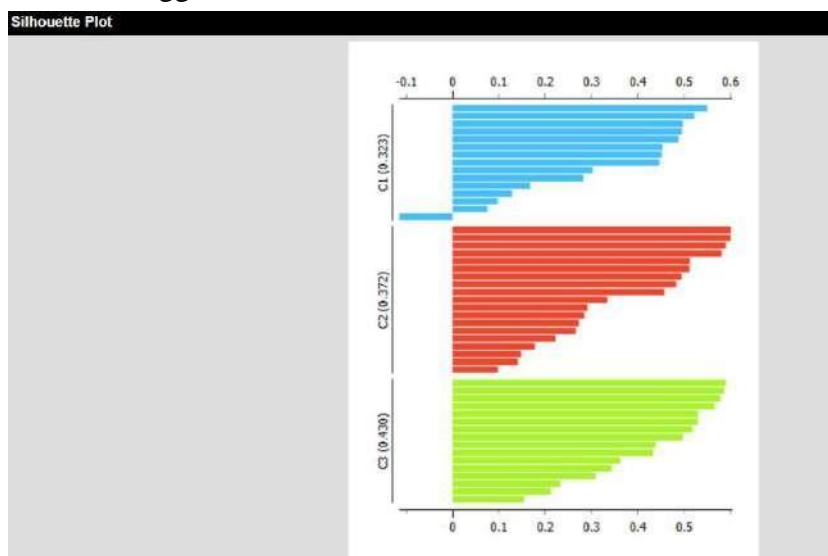
Gambar 4.16 Silhouette Plot K - Means

Pada gambar 3.16, hasil Silhouetter plot dengan menggunakan widget K – Means Orange Data Mining 0.021.



Gambar 4.17 PCA dan K - Means Orange

Pada gambar 4.17, di Orange Data Mining, penggunaan K-means dan PCA ( Principal Component Analysis ) dapat terintegrasi dalam aliran kerja analisis data. K-means cenderung bekerja lebih baik pada data yang memiliki dimensi yang lebih rendah. Dengan menggunakan PCA untuk mereduksi dimensi menjadi beberapa komponen, dapat meningkatkan kinerja K-means dan mengatasi masalah dimensi tinggi.



Gambar 4.18 Hasil K - Means dan PCA

Dengan menggunakan *silhouette plot*, dapat dipahami seberapa baik kluster – kluster tertentu sesuai dengan data, dan ini dapat membantu dalam pemilihan jumlah optimal kluster dalam algoritma pengelompokan seperti k-means. Semakin tinggi dan seragam nilai *silhouette*, semakin baik pembagian data ke dalam kluster.

PCA memungkinkan visualisasi data dalam dimensi yang lebih rendah. Dengan mereduksi dimensi data ke 2 atau 3, dapat divisualisasikan data dan *cluster* secara lebih efektif, membantu pemahaman tentang struktur dan pola dalam data. Hasil *Silhouette plot* seperti ditunjukkan sebelumnya 0.430 pada gambar 4.18.

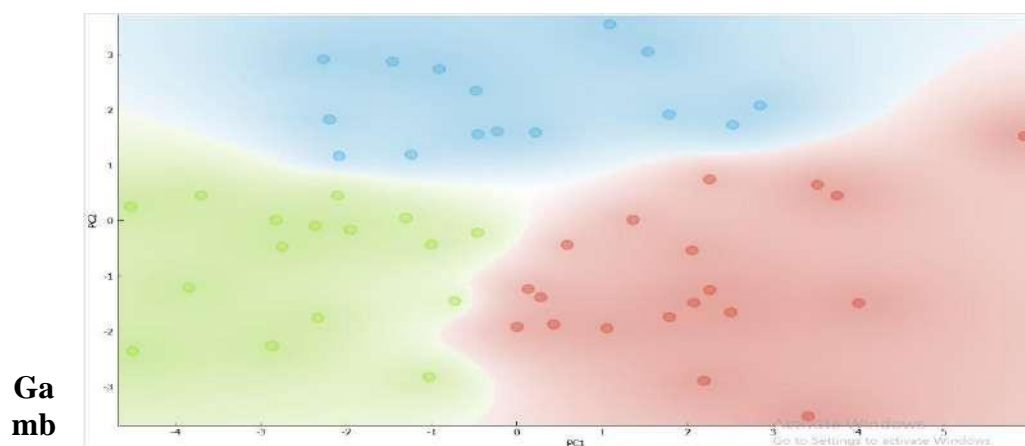
Silhouette Plot dengan K-Means	"Silhouette Plot PCA dan K - Means	Silhouette dengan <i>Script Python</i>
0.021	0.430	0,429748320637

Tabel 4.1 Tabel Perbandingan Silhouette

Dari tabel 4.1, hasil perbandingan *silhouette plot* dengan PCA dan K – Means memberikan hasil tingkat kesesuaian yang relatif baik dari objek-data terhadap kluster tempatnya berada dibandingkan dengan K – Means saja. Untuk *script* dengan *Python* pun juga memiliki kemiripan hasil.

#### 4.7 EVALUASI DENGAN ORANGE

Berdasarkan hasil penelitian dengan menggunakan Orange, hasil pengklasteran dengan gabungan PCA dan Orange



## ar 4.19 Hasil Scatter Plot

Pada gambar 4.19, hasil scatter plot menunjukkan bahwa kluster-klusternya terpisah dengan jelas di dalam diagram, yang berarti menunjukkan bahwa algoritma K-means berhasil memisahkan data menjadi kelompok-kelompok yang berbeda.

Selanjutnya, untuk mengetahui faktor yang paling berpengaruh terhadap rekomendasi profil lulusan SMK Negeri 6 Pekanbaru dapat ditunjukkan dengan menggunakan hasil PCA dengan penerapan nilai eigen. *Nilai eigen* dari matriks kovariansi atau *matriks korelasi* dapat memberikan informasi tentang seberapa tersebar variabilitas data dalam arah tertentu.

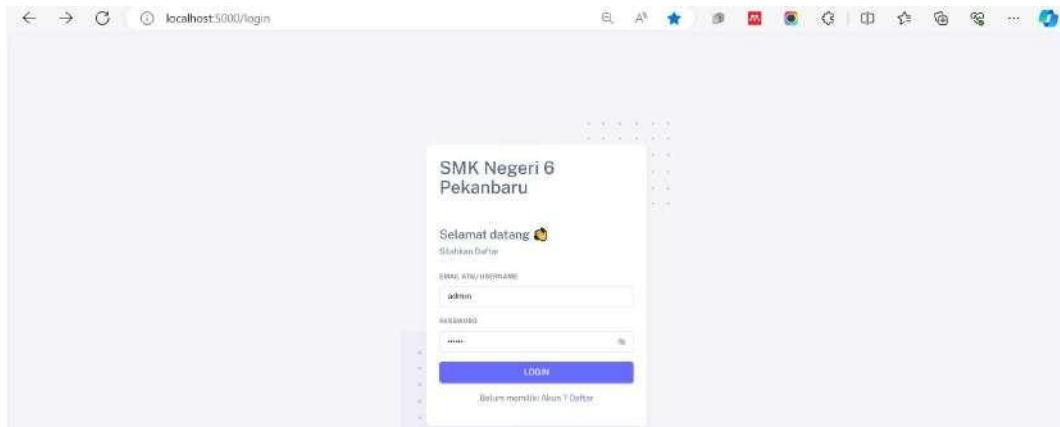
variance	Feature name	PC1 0.149094	PC2 0.0810481	alumni PC1	alumni PC2
5	Apakah Anda merupakan anak pertama?	0.0863776	0.269821	0.0863776	0.269821
6	Apakah Ayah Anda pernah atau sedang menjalankan sebuah usaha/ bisnis...	-0.0203475	0.192897	0.0203475	0.192897
7	Pendidikan terakhir Ayah?	-0.0607619	0.0974232	0.0607619	0.0974232
8	Apakah Ibu Saudara pernah atau sedang menjalankan sebuah usaha/ bisni...	-0.0110543	0.27092	0.0110543	0.27092
9	Pendidikan terakhir Ibu?	-0.0420107	0.00679916	0.0420107	0.00679916
10	Apakah Saudara pernah atau sedang menjalankan sebuah usaha/ bisnis sa...	0.0591467	0.164496	0.0591467	0.164496
11	Apakah Saudara aktif dalam organisasi Sekolah/ekstrakurikuler?	0.138159	0.106366	0.138159	0.106366
12	Apakah Saudara pernah mengikuti sebuah kursus saat sekolah di SMK?	-0.00755124	0.101017	0.00755124	0.101017
13	Apakah alasan masuk SMK?	0.0591685	-0.0714612	0.0591685	0.0714612
14	SMK tempat Saudara belajar menyediakan fasilitas yang lengkap?	0.0690402	-0.00427086	0.0690402	0.00427086
15	Apakah SMK tempat Saudara belajar menyediakan fasilitas yang up to dat...	0.130863	0.162169	0.130863	0.162169
16	Bagaimanakah biaya sekolah di SMK menurut Saudara saat masih berseko...	0.100775	0.0345473	0.100775	0.0345473
17	Apakah Saudara pernah mendapatkan beasiswa saat di SMK?	-0.058628	-0.112669	0.058628	0.112669
18	Apakah Saudara melanjutkan ke perguruan tinggi setelah lulus SMK?	-0.117792	-0.164391	0.117792	0.164391
19	Jika pertanyaan di atas adalah "Ya", pada jenjang pendidikan apakah Anda...	-0.123889	-0.209922	0.123889	0.209922
20	Jika ada yang bertanya, apakah Saudara akan merekomendasikan kepada ...	0.0342867	0.226692	0.0342867	0.226692
21	( Pertanyaan untuk Alumni ) Apakah materi yang diajarkan di sekolah ses...	-0.0404509	-0.0567064	0.0404509	0.0567064
22	( Pertanyaan untuk Alumni ) Apakah Saudara pernah atau sedang menjala...	0.077852	0.256913	0.077852	0.256913
23	( Pertanyaan untuk Alumni ) Apakah Saudara Bekerja saat ini?	0.0564404	0.14275	0.0564404	0.14275
24	( Pertanyaan untuk Alumni ) Apakah Bidang Pekerjaan Saudara saat ini ses...	-0.042141	0.0504906	0.042141	0.0504906
25	( Pertanyaan untuk Alumni ) Apakah SMK membantu Saudara dalam men...	0.0179254	0.203699	0.0179254	0.203699
26	PAI	-0.272407	0.0577984	0.272407	0.0577984
27	PKN	-0.266802	0.213266	0.266802	0.213266
28	B.INDO	-0.2683	0.133405	0.2683	0.133405
29	MTK	-0.270002	0.0305652	0.270002	0.0305652
30	SEJARAH	-0.182368	0.136518	0.182368	0.136518

Gambar 4.20 Hasil PCA

Pada gambar 4.20, *nilai eigen* yang lebih besar menunjukkan bahwa variabilitas data lebih dominan dalam arah tersebut. *Nilai eigen* value PCA1 adalah 0.149, dan *nilai eigen* value PCA2 adalah 0.081. Secara umum, *nilai eigen* value yang lebih besar sebanding dengan kontribusi komponen utama terhadap variasi data. *nilai eigen value total* adalah 1, dan *nilai eigenvalue kumulatif* PCA1 adalah 0.149 dan *nilai eigenvalue kumulatif* PCA2 adalah 0.081, maka varian

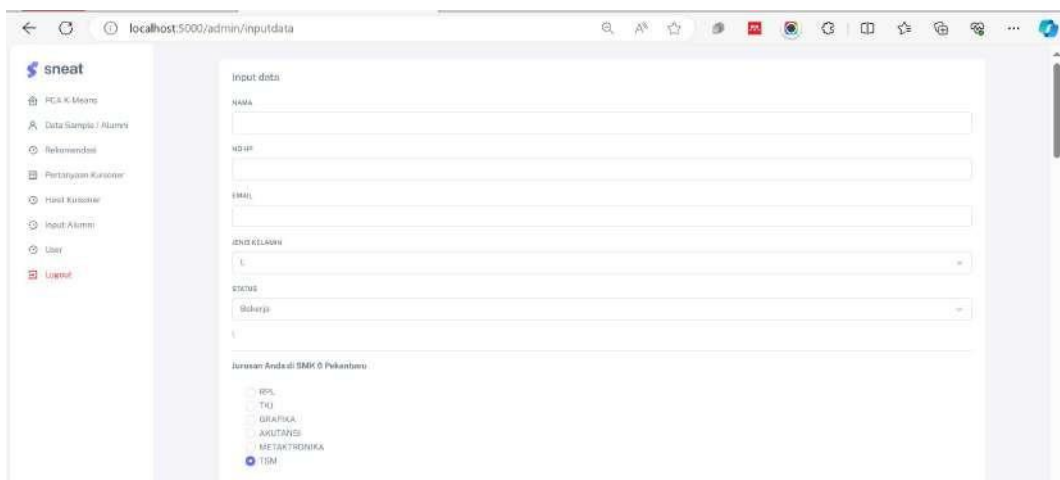
kumulatif yang dijelaskan oleh kedua komponen tersebut adalah  $0.149 + 0.081 = 0.230$ . Semakin tinggi nilai kumulatif *eigen value* yang mendekati 1, semakin banyak varians yang dijelaskan oleh komponen-komponen utama tersebut.

#### 4.8 PENGUJIAN SISTEM REKOMENDASI



Gambar 4.21 Halaman Login Pengujian

Pada gambar 4.21, terdapat halaman login untuk admin dan user untuk masuk ke dalam halaman sistem aplikasi.



Gambar 4.22 Halaman Input Kuisiener



Pada gambar 4.22, alumni dapat mengisi biodata terlebih dahulu sebelum menginput kuisioner.

variance	nama di SMK 6 Pk	Suku Anda?	kurikulum yang	dikembangkan	dari SMK tempo	Pada Tahun ber	Cluster	Silhouette	PC1 0.149094	PC2 0.081048
1	Ikhtik Komput...	-	Green community...	Lebih melengka...	Beberapa kali ...	2023	C1	0.657229	1.53389	3.05334
2	ekayasa peran...	Melayu	Seni suara	-	-	2022	C2	0.581499	0.430123	-1.87499
3	Pl	ndak tau	selma sekolah tl...	tingkat kan ola...	bnyk jamkos	2023	C1	0.656052	0.217555	1.59166
4	ekayasa Peran...	Minang	Club b Inggris	Tingkatkan lagi ...	Banyak sih,sala...	2021	C2	0.623709	3.41316	-3.53851
5	ekayasa Peran...	Minang	Paskibra atau P...	Cara ajar yang ...	Dekat dari rum...	2022	C3	0.563658	-0.460255	-0.219695
6	EKAYASA PER...	BATAK	PODCAST, GRE...	FASILITAS, TEN...	JURUSAN NYA ...	2023	C3	0.556888	-0.727873	-1.45535
7	ekayasa Peran...	Batak	Pasus	Kedisiplinan	Kreatif dan Disi...	2021	C1	0.6552	-0.231274	1.61055
8	PL	Minang	Pramuka	#NAME?	Memiliki ruang ...	<2021	C2	0.587183	3.75022	0.450875
9	ekayasa peran...	Jawa	Badminton	-	Memiliki labor s...	2023	C1	0.669467	-0.911892	2.73852
10	EKAYASA PER...	Jawa & Minang	-	#NAME?	Ramah lingkun...	2023	C1	0.673829	-0.480594	2.3477
11	CJ	Lahir asli pekan...	MUSIK	Moral dan etika...	Akreditasi sekol...	2021	C2	0.526535	5.94678	1.52719
12	ikhtik kompute...	Jawa	OSIS,passus	Fasilitas yang m...	Tempatnya bela...	2021	C2	0.567947	3.51918	0.645864
13	ekayasa peran...	Melayu	Tidak ada	Mengurangi pel...	Memiliki pemb...	2022	C1	0.661818	1.0084	3.54749
14	ekayasa Peran...	Jawa	Musik	Mempertingan l...	Banyak ilmu da...	2021	C1	0.657345	-1.45644	2.87274
15	ekayasa Peran...	Jawa	Marching band ...	Mengasah ilmu...	Jarak yang tida...	2022	C1	0.647876	-0.457582	1.55987
16	pl	Jawa	Tidak ada	Sistem pendidi...	Tempatnya dek...	2023	C3	0.656574	-2.36029	-0.097912
17	ekaya perangk...	minang	pramuka	jwa semangat ...	fasilitas cukup,p...	2023	C1	0.59052	-2.19589	1.82531
18	ekayasa Peran...	Jawa	tidak ada	menggratiskan ...	-	2023	C1	0.507767	-2.08472	1.16462
19	pl	Jambi	menari	menyediakan fa...	memiliki banya...	2022	C1	0.587149	2.84753	2.08157
20	ekayasa peran...	Suku Piliang (M...	Seni Tari	Sbg pendidik (g...	Perpustakaanrny...	2022	C1	0.578619	2.52923	1.7291
21	ikhtik Komput...	Minang	tidak ada	dengan mening...	adanya mentori...	2023	C2	0.562787	0.279037	-1.39037
22	CJ (TEKNIK KO...	-	-	Kurang tau	-	2023	C2	0.514701	2.25772	0.742327
23	ikhtik kompute...	Jawa	Pramuka	Belajar lebih tek...	Fasilitas lengkap	2023	C2	0.65827	4.00231	-1.49467
24	ekayasa Peran...	Melayu	English Club	Memperbanyak...	Guru guru yang...	2022	C2	0.532839	0.00409472	-1.92202

Gambar 4.23 Data PCA dan K-Means

Pada gambar 4.23, dengan menggunakan PCA, hasil *clustering* K – Means dapat lebih maksimal. Hasil *Silhouette* juga menunjukkan rentang antara 0.5 sampai 0.6. Dapat digambarkan setelah dilakukan pengujian dengan modelling menggunakan Python dan Orange Data Mining, didapat bahwa selain faktor fasilitas dan kurikulum yang selaras dengan industri, nilai pendidikan karakter berpengaruh besar pada rekomendasi profil lulusan SMK Negeri 6 Pekanbaru seperti nilai PKN, Seni Budaya, Agama dan Matematika.

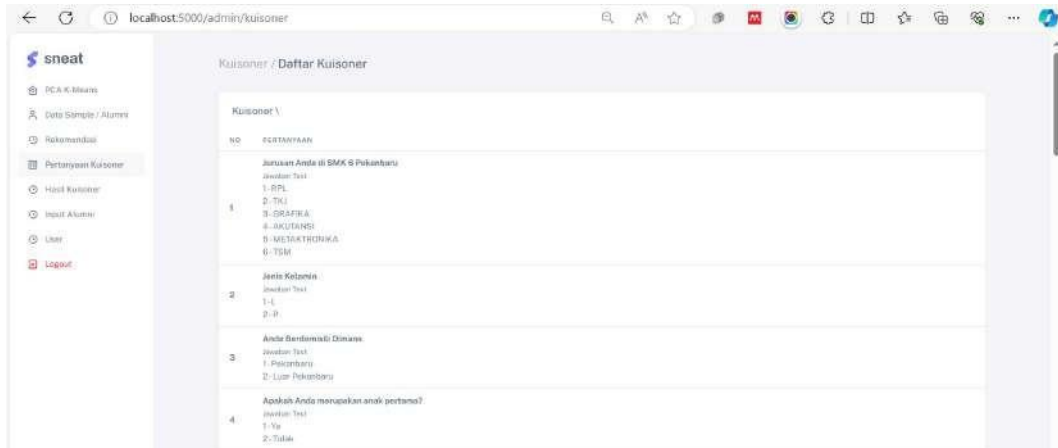
Gambar 4.24 Hasil Rekomendasi Sistem

Rekomendasi untuk sekolah agar dapat lulusannya dominan bekerja pada kluster 0 dengan persentase 38%, kluster 1 dengan persentase 32% melanjutkan pendidikan dan kluster 2 dengan persentase 30% bekerja dan wirausaha. Secara keseluruhan menurut alumni dari semua klaster, ditinjau dari rekomendasi penyediaan fasilitas terbaru berpendapat sama, masih kurang optimal, seperti keterbatasan ruang olahraga atau laboratorium tertentu. Kondisi ini mungkin mempengaruhi pengalaman peserta didik dalam mendapatkan pengajaran praktis dalam beberapa mata pelajaran. Begitu pula, untuk kelengkapan fasilitas sarana prasarana sudah terpenuhi, dan mendapatkan tanggapan positif dari alumni yang bekerja dan wirausaha, sedangkan dari segi kurikulum yang harus selaras dengan kurikulum industri, semua alumni menyatakan bahwa kurikulum di SMK Negeri 6 Pekanbaru didesain untuk mencakup aspek teori dan praktis. Adanya mata pelajaran yang relevan dengan dunia industri, penekanan pada keterampilan teknis, dan kerjasama dengan perusahaan memberikan peserta didik dasar yang kuat untuk memasuki dunia kerja, bahkan menurut alumni yang melanjutkan pendidikan, kurikulum yang disajikan sekolah sudah memenuhi dunia industri. Kurikulum juga berpengaruh pada bidang pekerjaan, semua alumni menyatakan sudah memenuhi. Dan menurut semua alumni baik dari semua kluster 0 yaitu bekerja, kluster 1 yaitu melanjutkan pendidikan dan kluster 2 yaitu bekerja sambil berwirausaha, informasi lowongan kerja yang disediakan oleh sekolah dirasa masih kurang dan perlu peningkatan. Dari hasil kuisioner alumni juga, perlu adanya peningkatan kompetensi lebih lanjut terutama terhadap mata pelajaran pengembangan pendidikan berkarakter ( soft skill ) sehingga lulusan dapat berhasil melakukan pencapaian.

No	Nama	Jenis Kelamin	Status	Pekerjaan	Aksi
1	Prillya Sulaksana	F	Bekerja dan Wirausaha		Lihat detail
2	Novian Rizki	F	Bekerja	Perawat	Lihat detail
3	Muhammad Rizki Pratiwi	F	Melanjutkan Pendidikan	Perawat	Lihat detail
4	Alvin Yudianto	F	Bekerja dan Wirausaha	Perawat	Lihat detail
5	Doni Febriani	F	Bekerja	Perawat	Lihat detail
6	KA RAHMADANI	F	Bekerja	Perawat	Lihat detail
7	Christa Marlene Ramona	F	Bekerja dan Wirausaha	Perawat	Lihat detail
8	Andriana Nugroho	F	Tidak Bekerja	Perawat	Lihat detail
9	MALINDA SARDIANI	F	Bekerja	Perawat	Lihat detail
10	Dipang Muhammad Zaki	F	Bekerja	Perawat	Lihat detail
11	MUHAMMAD HUSRI AL-SAMUDI	F	Bekerja	Perawat	Lihat detail
12	Rizka Damayanti	F	Melanjutkan Pendidikan	Perawat	Lihat detail
13	Fahmi Saadiah	F	Bekerja, Melanjutkan Pendidikan	Perawat	Lihat detail
14	Doni Adhika Veronika	F	Bekerja	Perawat	Lihat detail

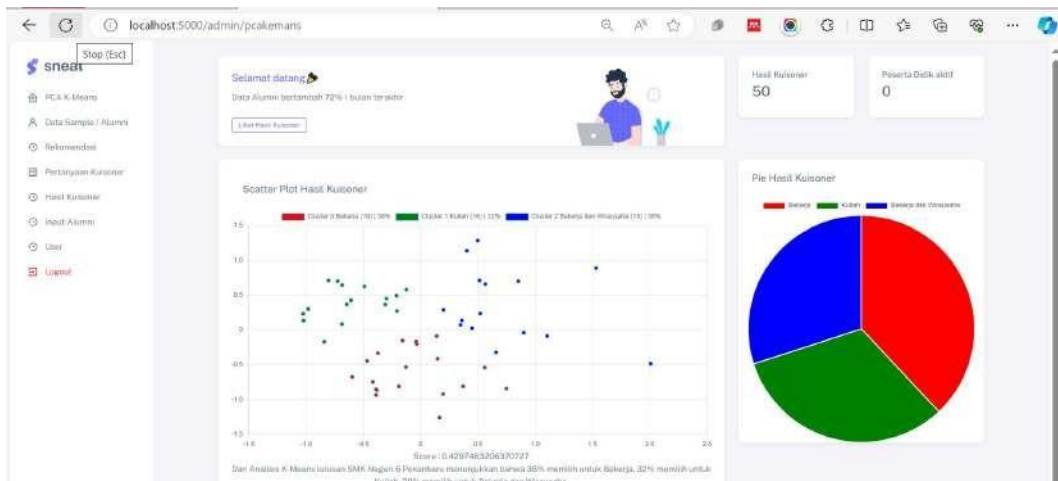
Gambar 4.25 Hasil Menu Daftar Alumni

Pada gambar 4.25, terdapat hasil menu Daftar Alumni yang terdiri atas nama, jenis kelamin, status, kontak, dan action “Lihat Data” yang hanya dapat dilihat pada halaman admin saja.



Gambar 4.26 Daftar Pertanyaan Kuisiner

Pada gambar 4.26, berisikan informasi daftar pertanyaan kuisiner yang dapat dilihat pada halaman user ataupun admin.



Gambar 4.27 Hasil Kuisiner

Pada gambar 4.27, terdapat hasil kuisiner dalam bentuk scatter plot dengan hasil menggunakan *script Python* 0.429748 dan diagram Pie dengan hasil dominan warna merah yaitu bekerja pada kluster 0 dengan jumlah alumni sebesar 38%, warna hijau yaitu melanjutkan pendidikan ( kuliah ) pada kluster 1 sebesar

32% dan warna biru yaitu bekerja sambil melanjutkan berwirausaha pada klaster 2 sebesar 30%.

Sebelum aplikasi ini dinyatakan selesai, penulis melakukan pengujian terlebih dahulu untuk melihat fungsi dari seluruh menu aplikasi berjalan dengan baik atau tidak. Pengujian ini menggunakan metode blackbox.

a. Pengujian Blackbox Validasi Login Admin ataupun User

No	Skenario Pengujian	Test Case	Hasil yang diharapkan	Hasil Pengujian	Kesimpulan
1.	Mengosongkan semua isian data pada form login, lalu mengklik tombol login	Username : - dan Password : -	Sistem akan menolak login dan akan menghasilkan pesan “Invalid Username dan Password”	Sesuai harapan	Valid
2.	Menginput dengan salah satu nilai benar baik username ataupun password lalu mengklik login.	Username : admin ( benar ) dan password ( salah ) ataupun sebaliknya. Hal ini berlaku juga pada halaman user.	Sistem akan menolak login dan akan menghasilkan pesan “Invalid Username dan Password”	Sesuai harapan	Valid
3.	Menginput dengan keduanya	Username : admin ( benar ) dan	Sistem masuk halaman admin apabila	Sesuai harapan	Valid

	benar baik username dan password	password ( benar ) Hal ini berlaku juga pada halaman user.	menggunakan login admin. Sedangkan, masuk ke halaman web user, apabila login user.		
--	--	--	--	--	--

Tabel 4.2 Pengujian Blackbox Login

b. Pengujian Blackbox Input User ataupun Admin

No	Skenario Pengujian	Hasil yang diharapkan	Hasil Pengujian	Kesimpulan
1.	Mengisi seluruh form isian baik pertanyaan ataupun nilai, lalu klik simpan	Sistem akan menyimpan data yang akan diinput ke dalam database.	Sesuai harapan	Valid

Tabel 4.3 Pengujian Input Alumni

## **BAB V**

### **KESIMPULAN DAN SARAN**

#### **5.1. KESIMPULAN**

Berdasarkan hasil penelitian mengenai Rekomendasi Profil Peserta Didik SMK dengan menerapkan metode machine learning menggunakan algoritma K-Means, ditemukan bahwa terdapat 3 klaster yang dapat menentukan arah karir para lulusan SMK. Data penelitian diperoleh melalui kuisioner kepada alumni dari kompetensi keahlian Rekayasa Perangkat Lunak dan Teknik Komputer Jaringan dan nilai. Data tersebut diolah menggunakan K-Means untuk memberikan rekomendasi kepada peserta didik kelas XII nantinya mengenai pilihan setelah tamat, seperti bekerja, melanjutkan pendidikan, atau berwirausaha. Hal ini diharapkan dapat membantu peserta didik membuat keputusan yang sesuai dengan minat, bakat, dan potensi mereka, menghasilkan lulusan SMK yang lebih terampil dan berkualitas. Dengan PCA, tahap visualisasi lebih muda dan analisis data dengan data yang sangat besar menjadi beberapa indeks yang lebih kecil lebih ringkas. Penerapan K-Means dan PCA diimplementasikan dalam aplikasi Python dengan menghasilkan *Silhouette Plot* mendekati 0,42974832, dengan mayoritas alumni memilih bekerja dibandingkan melanjutkan perguruan tinggi dan berwirausaha. Untuk memvalidasi hasil, aplikasi Orange juga digunakan dengan hasil yang lebih akurat, yaitu mendekati 0,430. Hasil penelitian menunjukkan bahwa nilai – nilai yang berfokus pada pendidikan berkarakter sangat mempengaruhi arah karir lulusan. Implementasi sistem ini diharapkan dapat membantu mengurangi kesenjangan antara keterampilan yang diajarkan di SMK dan kebutuhan dunia kerja.

#### **5.2 SARAN**

Berikut adalah beberapa saran yang dapat dibuat berdasarkan hasil penelitian:

1. Menurut penelitian, kualitas pembelajaran di SMK dipengaruhi oleh sarana dan fasilitas. Oleh karena itu, sekolah harus meningkatkan sarana dan prasarana mereka, seperti membangun laboratorium dan menyediakan peralatan modern yang sesuai dengan kebutuhan industri.
2. SMK disarankan terus mengembangkan dan memperbarui kurikulum mereka karena pentingnya program kurikulum yang sesuai dengan dunia industri. Berkolaborasi dengan perusahaan atau industri dapat membantu memastikan bahwa materi yang diajarkan sesuai dengan kebutuhan pasar kerja.
3. Bimbingan karir dapat membantu peserta didik kelas XII memahami pilihan karir setelah lulus.
4. Melihat bahwa sifat seperti PKN, Seni Budaya, dan Keterampilan interpersonal sangat memengaruhi jalan karir, disarankan agar SMK memberikan perhatian khusus pada pengembangan keterampilan soft skills seperti komunikasi, kerja tim, dan kepemimpinan.
5. SMK dapat membuat program penyusunan rencana pribadi untuk setiap peserta didik di kelas XII. Program ini dapat dibuat dengan mempertimbangkan minat, bakat, dan potensi masing-masing peserta didik, sehingga dapat memberikan panduan yang lebih personal untuk memilih arah karir setelah tamat sekolah.
6. Peserta didik dapat memperoleh pengalaman kerja praktis dengan mengikuti program magang atau pelatihan berbasis industri.
7. Disarankan agar sistem rekomendasi yang menggunakan machine learning secara berkala dievaluasi dan diperbarui karena dinamika dunia kerja yang terus berubah. Ini akan membantu memastikan bahwa rekomendasi yang diberikan selalu sesuai dengan kebutuhan dan tren terkini, apalagi dengan jumlah alumni yang bersedia dan mau mengisi input dari sistem rekomendasi terus bertambah, sehingga data makin valid

## DAFTAR PUSTAKA

( 无 日 期 ). 检 索 来 源 :

<https://pusatinformasi.raporpendidikan.kemdikbud.go.id/hc/en-us/articles/10446868880537-Mengunduh-Perencanaan-Berbasis-Data-PBD-untuk-Pendidikan-Dasar-dan-Menengah-Dikdasmen->

<https://pusatinformasi.raporpendidikan.kemdikbud.go.id/hc/en-us/articles/10446868880537-Mengunduh-Perencanaan-Berbasis-Data-PBD-untuk-Pendidikan-Dasar-dan-Menengah-Dikdasmen->

1. Ariyani LFSUM, Wahyono H, Haryono A, Rusdi JF, Pratama CBA. Vocational education phenomena research method. MethodsX. 2021Widjaja, & 8. (无日期).

(2022). 检 索 来 源 : <https://www.bps.go.id:https://www.bps.go.id/indicator/6/674/1/pengangguran-terbuka-menurut-pendidikan-tertinggi-yang-ditamatkan.html>

(2022

AfriyaniOktaNova. ( 无 日 期 ). Peranan Keluarga Terhadap Prestasi Siswa Pada Mata Pelajaran Ilmu Pengetahuan Sosial.

AlmahyIbrahim. (2013). Database integration: Importance and approaches.

Journal of Theoretical and Applied Information Technology, 54, 150-154.

AzizahNurHanifah. (2023). Rstudio. 检 索 来 源 : Rpubs:  
[https://rpubs.com/hannnnnnnnnf\\_/1109252](https://rpubs.com/hannnnnnnnnf_/1109252)



- AzraB. (2021). the Development of Engineering Drawing E-Module for Grade X At Smk (Vocational School). *Jurnal Teknologi Informasi dan Pendidikan*(460).
- BrownleeJason. (2017). Machine Learning Mastery". 检索来源 : <https://machinelearningmastery.com>
- DanaDanarRaditya. (2021). Competency test clustering through the application of Principal Component Analysis (PCA) and the K-Means algorithm. *IOP Conference Series: Materials Science and Engineering*, 012038.
- Data)PRINCIPAL COMPONENT ANALYSIS (PCA) SEBAGAI PENANGANAN ASUMSI MULTIKOLINEARITAS (STUDI KASUS: DATA PRODUKSI TAPIOKA) Principal Component Analysis (PCA) Method to Handle Multicollinearity Assumptions (Case Study: Tapioca Production)METODE. (2023). Retno, Dwi. 115-124.
- FikriRijalul. (2021). Pengelompokan Kabupaten/Kota di Indonesia Berdasarkan Informasi Kemiskinan Tahun 2020 Menggunakan Metode K-Means Clustering Analysis. *Seminar Nasional Teknik dan Manajemen Industri*, 190-199.
- GustientiedinaGustientiedina. (2019). Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan. *Jurnal Nasional Teknologi dan Sistem Informasi*, 17-24.
- Halasan Simanullang1)Ari Sapto3)Wahjoedi2),. (无日期). Peran Lingkungan Keluarga Dalam Meningkatkan Prestasi Belajar Siswa.
- Hemmati-SarpardehAbdolhossein. (2020). Chapter 1 - Introduction.
- JohnsonMooreSusan. (2008). The next generation of teachers: Who enters, who stays, and why.
- KuritaTakio. (2019). *Principal Component Analysis (PCA) BT - Computer Vision: A Reference Guide*. Champ: Springer International Publishing.

- MarukatSanparith. (2023). Tutorial on PCA and approximate PCA and approximate kernel PCA. Springer Netherlands.
- Mukhidin MukhidinMupitaJonah. (2019). Factors That Affect the Quality Status of Vocational High Schools. Proceedings of the 5th UPI International Conference on Technical and Vocational Education and Training (ICTVET 2018). Atlantis Press.
- NasutionZulfahmiMuhammad. (2019). PENERAPAN PRINCIPAL COMPONENT ANALYSIS (PCA) DALAM PENENTUAN FAKTOR DOMINAN YANG MEMPENGARUHI PRESTASI BELAJAR SISWA (Studi Kasus : SMK Raksana 2 Medan). Jurnal Teknologi Informasi, 41.
- PutraFirmansyahSepta. (2016). Feature Selection pada Dataset Faktor Kesiapan Bencana pada Provinsi di Indonesia Menggunakan Metode PCA (Principal Component Analysis). urnal Teknik ITS.
- RetnoDwi. (2023). METODE PRINCIPAL COMPONENT ANALYSIS (PCA) SEBAGAI PENANGANAN ASUMSI MULTIKOLINEARITAS (STUDI KASUS: DATA PRODUKSI TAPIOKA) Principal Component Analysis (PCA) Method to Handle Multicollinearity Assumptions (Case Study: Tapioca Production Data). 115-124.
- SurohmanSurohman. (2021). Korelasi Antara Profil dan Nilai Akademis Siswa dengan Menggunakan Algoritma K-Means. Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK) DOI: 10.25126/jtiik.202183034, 845.
- YuliantiRistiTitis. (2022). Identifikasi Atribut dengan Principal Component Analysis dan K-Means Clustering Sebagai Dasar Penyusunan Strategi Promosi KB Pria di Indonesia. Jurnal Biostatistik, Kependudukan, dan Informatika Kesehatan, 79.

