

Beberapa bagian dalam dokumen tesis ini telah disunting untuk menghapus informasi pribadi atau sensitif, seperti nama pembimbing, nama universitas, dan data pribadi lainnya. Penyuntingan dilakukan guna menjaga privasi serta kepatuhan terhadap etika berbagi dokumen akademik di ruang publik.

Gerry Aria Rotama

## INTISARI

Di era saat ini banyak masalah yang timbul dikalangan mahasiswa diberbagai universitas di negara kita, baik itu salah memilih jurusan atau lama dalam menyelesaikan studi yang diemban selama perkuliahan. Salah satu faktor yang menyebabkan lamanya mahasiswa menyelesaikan suatu perkuliahan adalah salah dalam menentukan Tugas Akhir apa yang akan diambil, sehingga sering terjadi pergantian judul ditengan penelitian, atau bahkan lamanya pengerjaan penelitian itu sendiri. Atas dasar itu penulis melihat dalam suatu pengambilan keputusan dibutuhkan faktor yang penting dalam standar untuk mengabil keputusan itu sendiri. Dalam penelitian ini sendiri, pemilihan atribut yang akan digunakan sudah tepat dalam menentukan standar dari hasil yang akan menjadi patokan dari rekomendasi judul tugas akhir dan juga input yang akan menghasilkan variabel data pun sudah tepat sasaran sehingga dapat menghasilkan data yang akurat. Yang mana penelitian ini nanti akan memiliki pengaruh yang signifikan dalam menaikkan kualitas jurusan, karena dapat meningkatkan jumlah mahasiswa yang lulus tepat waktu, dimana jika mahasiswa mengambil tugas akhir yang sesuai dengan bidang yang dikuasai maka akan mempercepat pengerjaan tugas akhir. Terkait metode yang akan digunakan, penulis memilih untuk menggunakan *Support Vector Machine (SVM)* hal ini bukan tanpa alasan, penulis memilih *SVM* sebagai metode yang akan digunakan dikarenakan metode ini sendiri memilik keunggulan tersendiri pada atribut yang akan penulis gunakan. Dimana *SVM* dapat digunakan untuk klasifikasi dan regresi dimana hal ini sangat sesuai untuk menentukan Tugas Akhir yang tepat untuk Mahasiswa/i. Dan dari hasil penelitian yang sudah didapatkan, bisa diamati bahwa data yang tidak penggunaan *SMOTE* memiliki akurasi yang jauh lebih kecil dan juga diikuti dengan prediksi, *class recall*, dan *class precision* yang tidak seimbang, hal ini disebabkan oleh data yang inbalance dan juga faktor matakuliah yang berkaitan dengan label yang tidak seimbang. Sehingga dibutuhkan *SMOTE* untuk meningkatkan akurasi dan menyeimbangkan data, meskipun tidak dapat mempengaruhi matakuliah yang tidak seimbang terhadap ke 4 label. Dan dari penelitian yang penulis lakukan, dapat disimpulkan bahwa penggunaan *SVM* dengan operator *SMOTE Upsampling* yang disertakan *Shuffled Sampling* adalah pengaturan yang memiliki akurasi paling tinggi dibandingkan dengan pengaturan lain.

**Kata kunci** ; Rekomendasi, Tugas Akhir, Mahasiswa, Data Mining, SVM, SMOTE Upsampling, Imbalance data

## ABSTRACT

*In the current era, there are many problems that arise among students at various universities in our country, whether it's choosing the wrong major or taking a long time to complete the studies carried out during lectures. One of the factors that causes students to take a long time to complete a course is wrong in determining what Final Project to take, so that there is often a change of title in the middle of research, or even the length of time the research is being carried out. On that basis, the authors see that in making a decision, an important factor is needed in the standard for making the decision itself. In this study itself, the selection of the attributes to be used was appropriate in determining the standards of the results that would become the benchmark for the final project title recommendations and also the inputs that would produce data variables were right on target so as to produce accurate data. Which this research will later have a significant influence on in improving the quality of majors, because it can increase the number of students who graduate on time, where if students take the final project in accordance with the field mastered will speed up the completion of the final project. Regarding the method that will be used, the author chooses to use a Support Vector Machine (SVM) this is not without reason, the author chooses SVM as the method to be used because this method itself has its own advantages in the attributes that the author will use. Where SVM can be used for classification and regression where this is very suitable for determining the right final project for students. And from the research results that have been obtained, it can be observed that data that does not use SMOTE has much smaller accuracy and is also followed by unbalanced predictions, class recall, and class precision, this is caused by imbalanced data and also subject factors that associated with unbalanced labels. So that SMOTE is needed to improve accuracy and balance the data, although it cannot affect courses that are not balanced against the 4 labels. And from the research that the authors conducted, it can be concluded that the use of SVM with the SMOTE Upsampling operator which includes Shuffled Sampling is the setting that has the highest accuracy compared to other settings.*

**Keywords ;** *Recommendations, Final Projects, Students, Data Mining, SVM, SMOTE Upsampling, inbalance data.*

## DAFTAR ISI

HALAMAN PENGESAHAN I.....	ii
HALAMAN PENGESAHAN II.....	iii
KATA PENGANTAR .....	iv
INTISARI .....	vi
ABSTRACT.....	viii
DAFTAR ISI.....	x
DAFTAR GAMBAR .....	xi
DAFTAR TABEL.....	xii
BAB I.....	9
1.1    Latar Belakang .....	9
1.2    Rumusan Masalah .....	9
1.3    Tujuan Penelitian.....	10
1.4    Sistematika Penulisan .....	10
BAB II.....	11
2.1    Tinjauan Pustaka .....	11
2.2    Landasan Teori .....	12
BAB III .....	17
3.1    Alat dan Bahan.....	17
3.2    Tahapan Penelitian .....	18
3.3    Studi Literatur .....	18
3.4    Pengumpulan Data .....	18
3.5    Pre-Processing Data.....	19
3.6    Implementasi Algoritma.....	23
3.7    Analisis Hasil.....	29
BAB IV .....	30
4.1    Implementasi Algoritma.....	30
4.2    Analisis Proses dan Hasil Pengujian .....	34
BAB V .....	35
5.1    Kesimpulan.....	35
5.2    Saran.....	35
DAFTAR PUSTAKA .....	36

## DAFTAR GAMBAR

Gambar 1. 1 Curva Sederhana Support Vector Machine.....	13
Gambar 3. 1 Flowchart .....	18
Gambar 3. 2 data cleaning .....	20
<i>Gambar 4. 3 Proses Import Excel</i> .....	25
Gambar 4. 4 Operator SMOTE Upsampling .....	26
Gambar 4. 5 Menambahkan Nominal to Numerical .....	27
Gambar 4. 6 Menambahkan Operator Cross Validation.....	27
Gambar 4.7 melakukan setting cross validation .....	28
Gambar 4. 8 pengaturan parameter .....	29
Gambar 4. 9 run rapidminer.....	29

## DAFTAR TABEL

Tabel 3. 1 data awal yang dikumpulkan.....	19
Tabel 3.2 data selection.....	20
Tabel 3. 3 transformasi atribut matakuliah .....	21
Tabel 3. 4 Atribut Label.....	22
Tabel 4. 1 Hasil Rapidminer dengan 1 operator SMOTE Upsampling .....	30
Tabel 4. 2 Hasil Rapidminer dengan 3 operator SMOTE Upsampling .....	30
Tabel 4. 3 Rumus Perhitungan Precision, Akurasi dan Recall.....	31
Tabel 4. 4 hasil pengujian sampling dan smote .....	32
Tabel 4. 5 Kesalahan Prediksi Label Data .....	33
Tabel 4. 6 Tabel Kesalahan Prediksi Label UI/UX.....	33



# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Faktor yang menyebabkan lamanya mahasiswa menyelesaikan suatu perkuliahan adalah salah dalam menentukan Tugas Akhir apa yang akan diambil, sehingga sering terjadi pergantian judul ditengan penelitian, yang mengakibatkan lamanya pengerjaan penelitian. Hasil dari tugas akhir yang dikerjakan menjadi salah satu kunci yang dapat menentukan kesuksesan dalam mengikuti perkuliahan disuatu program studi (Nurul Renaningtias, 2021). Selain itu pemilihan *Data Mining* dalam melakukan penelitian ini bukan tanpa alasan, hal ini dikarenakan *Data Mining* dapat digunakan untuk mengambil keputusan dimasa depan (Fitri Pratiwi, 2019). Sehingga hasil dari data yang didapatkan dari penelitian dapat menghasilkan rekomendasi yang akurat dan dapat mencapai target dari penelitian itu sendiri. Terkait metode yang akan digunakan, penulis memilih untuk menggunakan *Support Vector Machine (SVM)* dikarenakan metode dapat digunakan untuk klasifikasi dan regresi dimana hal ini sangat sesuai untuk menentukan Tugas Akhir yang tepat untuk Mahasiswa/i.

Penelitian ini akan memiliki pengaruh yang signifikan dalam menaikkan kualitas jurusan, karena dapat meningkatkan jumlah mahasiswa yang lulus tepat waktu, dimana jika mahasiswa mengambil tugas akhir yang sesuai dengan bidang yang dikuasai maka akan mempercepat pengerjaan tugas akhir. Maka dari itu penulis melakukan penelitian ini, selain dikarenakan memiliki pengaruh penting terhadap kualitas jurusan, hal ini juga dikarenakan topik dari penelitian yang dilakukan penulis belum pernah dilakukan sebelumnya, yang mana penulis menggunakan metode SVM dalam melakukan pengolahan data, yang mana seperti yang sudah dijelaskan di atas SVM sangat cocok dengan kasus yang akan diteliti oleh penulis.

### **1.2 Rumusan Masalah**

Berdasarkan latar belakang di atas, penulis akan melihat apakah metode SVM dapat melakukan prediksi yang akurat dalam terhadap rekomendasi tugas akhir. Dan juga apakah *SMOTE Upsampling* memiliki akurasi yang lebih baik.

Selain itu penulis juga menyimpulkan penelitian sebelumnya tidak memiliki atribut yang sama seperti penelitian yang dilakukan penulis

### **1.3 Tujuan Penelitian**

Mengimplementasikan metode *Support Vector Machine* untuk rekomendasi tugas akhir mahasiswa teknologi informasi.

### **1.4 Sistematika Penulisan**

Sistematika dari penulisan skripsi ini adalah sebagai berikut :

## **BAB I PENDAHULUAN**

Bab I menjelaskan tentang pendahuluan yang meliputi latar belakang permasalahan, rumusan masalah, tujuan, dan sistematika penulisan pada tugas akhir ini.

## **BAB II LANDASAN TEORI**

Bab ini menjelaskan tinjauan pustaka dan teori-teori pemecah masalah yang digunakan sebagai pendukung segala sesuatu yang berhubungan dengan topik penelitian ini.

## **BAB III METODOLOGI PENELITIAN**

Bab ini menjelaskan tentang metodologi yang digunakan pada penelitian dan membuat rancangan sistem agar dapat diimplementasikan sesuai harapan dengan mengacu mengacu pada teori-teori penunjang dan metode yang sudah dijelaskan pada bab sebelumnya.

## **BAB IV HASIL DAN PEMBAHASAN**

Bab IV menjelaskan hasil yang diperoleh dari seluruh penelitian dan dilakukan pengujian terhadap hasil implementasi sistem kemudian menganalisa agar sistem berjalansesuai dengan perancangan pada bab-bab sebelumnya.

## **BAB V PENUTUP**

Pada bab ini merupakan penutup yang berisi kesimpulan dan saran yang diambil dari penelitian ini dan saran untuk pengembangan lebih lanjut.

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 Tinjauan Pustaka**

Tugas akhir adalah kewajiban akhir yang akan dilaksanakan oleh mahasiswa dalam dunia perkuliahan, dimana hal ini dapat menjadi patokan untuk melihat seberapa dalam mahasiswa memahami mata kuliah yang telah diberikan selama perkuliahan, dimana nantinya mahasiswa akan memilih untuk fokus mengambil beberapa pilihan judul yang biasanya berupa saran dari dosen ataupun ide dari mahasiswa itu sendiri, hal ini sangat berperan penting dalam lulus atau tidaknya mahasiswa, dimana pengerjaan tugas akhir ini membutuhkan pemahaman akan mata kuliah yang berkaitan dengan judul tugas akhir yang akan diambil, sehingga mahasiswa harus mengetahui kemampuan dan pemahaman terhadap materi ataupun matakuliah yang akan menjadi landasan dari tugas akhir itu sendiri. Dikarenakan belum ada penelitian untuk merekomendasikan Tugas Akhir yang akan diambil mahasiswa, khususnya di jurusan teknologi informasi, maka penulis akan memaparkan beberapa penelitian yang dilakukan dengan konsep *data minning* yang berkaitan dengan dunia perkuliahan.

Dalam penelitian yang dilakukan oleh (Fitri Pratiwi, 2019) tentang analisis perancangan Algoritma Fuzzy C-Means dalam menentukan dosen pembimbing tugas akhir, dimana Algoritma fuzzy c-means merupakan suatu Teknik pengelompokan data dimana keberadaan tiap titik data dalam suatu cluster ditentukan oleh derajat keanggotaan. Fuzzy c-means adalah algoritma pengelompokan yang terawasi, karena pada algoritma fuzzy c-means jumlah cluster yang akan dibentuk perlu diketahui terlebih dahulu. Konsep dasar algoritma fuzzy c-means adalah menentukan pusat kelompok yang akan menandai lokasi rata-rata untuk tiap-tiap cluster. Penelitian ini bertujuan untuk menentukan dosen pembimbing sesuai dengan minat dan bidang keahlian dosen dengan metode clustering algoritma fuzzy c-means. FCM memiliki kelemahan sensitivitas terhadap nilai awal, mudah jatuh ke solusi optimal lokal dan sensitivitas terhadap

noise (Nameirakpam Dhanachandra, 2020). Manfaat yang diharapkan dari penelitian ini adalah agar pembagian dosen pembimbing tugas akhir merata sesuai dengan kepangkatan dan bidang keahlian masing-masing dosen. Namun penelitian ini hanya sebatas perancangan perhitungan menggunakan algoritma fuzzy c-means dan belum sampai ke tahap penelitian, hal ini dikarenakan syarat penentuan dosen pembimbing hanya berdasarkan kepangkatan, jumlah penelitian dan matakuliah yang pada bidang keahliannya yang mana algoritma fuzzy c-means itu sendiri adalah salah satu algoritma dalam metode clustering yang sangat baik dalam melakukan clusterisasi terhadap data dalam kondisi high level, yang mana hal ini membutuhkan jumlah variabel data yang banyak.

Tepat dalam memilih tugas akhir sangat berpengaruh terhadap lama atau tidaknya waktu yang dibutuhkan mahasiswa untuk lulus dari perkuliahan dan melaksanakan wisuda yang mana, hal ini juga akan berpengaruh terhadap akreditasi jurusan maupun kampus yang meluluskan mahasiswa tersebut. Oleh karena itu diperlukan rekomendasi yang tepat terhadap tugas akhir yang akan diambil oleh mahasiswa itu sendiri, sehingga hal ini akan membantu mahasiswa mengerjakan tugas akhir lebih cepat. Berdasarkan kondisi tersebut Teknik data minning akan sangat membantu dalam merekomendasikan tugas akhir itu sendiri, yang mana penelitian ini sendiri akan menggunakan kumpulan data data dari beberapa mahasiswa yang sudah menempuh Pendidikan dari smester 1 sama smester 6.

## **2.2 Landasan Teori**

### **2.2.1 Data Minning**

Menurut (Sagala, 2021) kata mining merupakan kiasan dari bahasa inggris, mine. Jika mine berarti menambang sumber daya yang tersembunyi didalam tanah, maka Data mining merupakan penggalian makna yang tersembunyi dari kumpulan data. Berikut ini tahapan proses data mining:

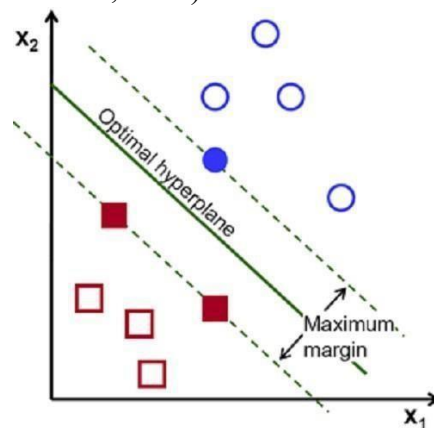
1. Pembersihan Data yaitu: menghapus data pengganggu (noise) dan mengisi data yang hilang.
2. Integrasi Data yaitu: menggabungkan berbagai sumber data.

3. Pemilihan Data yaitu: memilih data yang relevan.
4. Transformasi Data yaitu: mentransformasi data ke dalam format untuk diproses dalam data mining.
5. Proses mining yaitu: Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.
6. Evaluasi pola yaitu: berguna untuk mengidentifikasi pola-pola kedalam knowledge based yang ditemukan.

### 2.2.2 *Support Vector Machine (SVM)*

Menurut (Auliya Rahman Isnain, 2021) Support Vector Machine merupakan sistem pembelajaran yang menggunakan hipotesis berupa fungsi-fungsi linear dalam sebuah fitur yang berdimensi tinggi dan dilatih dengan menggunakan algoritma pembelajaran yang didasarkan dengan teori optimasi, Support Vector Machine diperkenalkan pertama kali oleh Vapnik pada tahun 1992 sebagai rangkaian harmonis konsep konsep unggulan dalam bidang pattern recognition, Tingkat akurasi pada model yang akan dihasilkan oleh proses peralihan pada svm sangat bergantung pada fungsi kernel dan parameter yang digunakan.

Secara sederhana SVM memiliki konsep mencari hyperplane“terbaik” yang berfungsi sebagai batas dari dua buah class. SVM mencari ini hyperplane berdasarkan support vector dan margin. Support vector adalah titik–titik yang paling dekat dengan separating hyperplane sedangkan margin menyatakan lebar dari separating hyperplane (Jumeilah, 2017).



**Gambar 1. 1 Curva Sederhana Support Vector Machine**

Pada Gambar 1.1 memperlihatkan beberapa data yang merupakan anggota dari dua buah kelas disimbolkan dengan kotak (merah) dan lingkaran (biru). Garis hijau pada Gambar 1.1 adalah yang disebut sebagai Hyperplane dan jarak antara garis putus-putus dan Hyperplane disebut fungsi pemisah. Margin sendiri adalah jarak antara fungsi pemisah tersebut dengan data terdekat dari masing-masing kelas dan data yang paling dekat inilah yang disebut sebagai *Support Vector*. Klasifikasi pada proses text mining dapat dikerjakan dengan menggunakan berbagai jenis metode klasifikasi yang salah satunya yaitu metode SVM (Debby Alita, 2020)

### 2.2.3 Rapid Miner

RapidMiner adalah aplikasi yang digunakan untuk menganalisis data secara kuantitas dan kualitatif untuk mendapatkan informasi dan pengetahuan sesuai dengan yang diharapkan (Muhammad Zamroni Uska, 2020). Dalam penggunaan Machine Learning pasti selalu dibutuhkan software pendukung. Untuk data mining sendiri memiliki banyak software yang dapat membantu pengerjaan dalam mengolah data di antaranya ada Rapid Miner, weka, clementine, tanagra dan lain-lain. RapidMiner merupakan perangkat lunak yang bersifat terbuka (*open source*). RapidMiner adalah sebuah solusi untuk melakukan analisis terhadap *data mining*, *text mining* dan analisis prediksi. RapidMiner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik.

Adapun kekurangan RapidMiner yaitu tidak bisa menyimpan hasil dari proses data mining yang dilakukan, yang bisa disimpan dalam RapidMiner hanyalah prosesnya saja. Penyimpanan proses juga hanya bisa di *local repository* yang ada dalam RapidMiner, bisa juga melakukan *export* dan *import* proses melalui RapidMiner tetapi proses yang di *export* akan berformat *.rmp* (*rapid miner process*) begitu juga dengan proses yang di *import* harus berformat *.rmp* juga. File berformat *.rmp* hanya bisa dibuka melalui rapidminer dengan perintah *import process* dibagian *file*.

### 2.2.4 Multi Class SVM

*Multi-class SVM* menyelesaikan masalah yang lebih dari dua kelas karena *SVM* hanya dapat melakukan klasifikasi dua kelas (biner). *SVM* berbeda dengan metode-metode lainnya yang dapat melakukan klasifikasi lebih dari dua kelas secara langsung. keakuratan *SVM* multi-kelas dilakukan dengan bantuan kernel linier fungsi dengan 500 iterasi (Utpal Barman, 2020). Kemudian *SVM* mengembangkan metode untuk melakukan multiclass, Terdapat 3 pendekatan yang dimiliki oleh *Multi-class SVM* yaitu *one-against-all(OAA)*, *one-against-one(OAO)* dan *error correcting output code*. Ada beberapa istilah dalam *classification performance* seperti *accuracy*, *class precision* dan *class recall*. *Accuracy* merupakan rasio prediksi Benar (positif dan negatif) dengan keseluruhan data. *Class recall* merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif. Sedangkan *class precision* merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif. Dalam klasifikasi *multi-class* rumus untuk mencari *accuracy*, *class recall* dan *class precision* dapat dilihat pada tabel 2.1 sebagai berikut :

**Table 2.1 dua Multi-class SVM**

	Kelas A	Kelas B	Kelas C
Pred A	VA1	VA2	VA3
Pred B	VB1	VB2	VB3
Pred C	VC1	VC2	VC3

$$Accuracy = \frac{(VA1 + VB2 + VC3)}{VA2 + VA3 + VB1 + VB3 + VC1 + VC2 + VA1 + VB2 + VC3}$$

$$Class Recall KELAS A = (VA1 + CB1 + VC1)$$

$$Class Precision KELAS A = \frac{(VA1)}{VA1 + VA2 + VA3}$$

#### 2.2.5 LBSVM

LIBSVM atau Library SVM, adalah sebuah Library pembelajaran komputer yang dikembangkan oleh Chih-Chung Chang dan Chih-jen Lin dari Universitas Nasional Taiwan pada tahun 2009. LIBSVM ini mempunyai 2 fungsi, yaitu fungsi Train dan Predict (Fathurrohman, 2020), selain itu fungsi dari LIBSVM adalah

sebagai open source machine learning library yang populer, dikembangkan di Universitas Nasional Taiwan dan keduanya ditulis dalam C++ meskipun dengan C API. LIBSVM mengimplementasikan algoritma Sequential minimal optimization (SMO) untuk mesin vektor dukungan kernelized (SVM), mendukung klasifikasi dan regresi. LIBSVM bisa digunakan untuk klasifikasi dengan label multiclass.

#### 2.2.6 SMOTE Upsampling

SMOTE merupakan salah satu turunan teknik over-sampling untuk menanggulangi ketidakseimbangan kelas dengan menyeimbangkan dataset dengan meningkatkan ukuran kelas minor (Femi Dwi Astuti, 2021). Penggunaan SMOTE pada penelitian ini berfungsi untuk mengantisipasi data yang tidak seimbang, dimana data yang tidak seimbang akan memiliki pengaruh kedalam akurasi yang dihasilkan didalam *RapidMiner*, sehingga dibutuhkan penyeimbangan data. Penggunaan SMOTE



## **BAB III METODOLOGI**

### **3.1 Alat dan Bahan**

Dalam melakukan penelitian ini penulis membutuhkan alat dan bahan sebagai penunjang untuk menyelesaikan penelitian dengan baik, yang akan dijelaskan sebagai berikut :

#### **3.2.1 Alat**

Dibawah ini merupakan spesifikasi dari perangkat keras (*hardware*) yang akan digunakan dalam penelitian dan penulisan, antara lain sebagai berikut:

- a. *Intel Core i*
- b. *DDR4 8 GB RAM*
- c. *Hardisk 1 TB*

Berikut ini adalah spesifikasi perangkat lunak (*software*) Laptop yang penulis gunakan untuk melakukan penelitian dan penulisan :

- a. *Operating System windows 64-bit*
- b. *Microsoft Excel 365*
- c. *RapidMiner Studio 9.3*

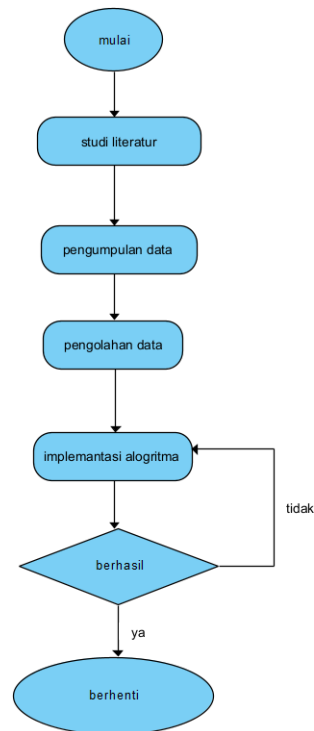
#### **3.2.2 Bahan**

Dalam penelitian dan penulisan, peneliti mengumpulkan data bahan dari TU Jurusan Teknologi Informasi yang berupa data nilai mahasiswa dan judul tugas akhir Angkatan 2016-2018 di jurusan Teknologi Informasi, yang mana dapat dilihat di bawah ini:

- a. Nim
- b. Nilai
- c. Judul ta
- d. Nama Mahasiswa
- e. IPK
- f. Kode Matakuliah
- g. Kelas
- h. Dosen Pembimbing
- i. Dosen Pembimbing 2
- j. Nilai TA
- k. Dosen Penguji
- l. Nama Matakuliah

### 3.2 Tahapan Penelitian

Dalam sebuah penelitian tentunya diperlukan tahapan yang jelas dalam mengerjakan, sehingga eksekusi yang akan dilakukan dapat lebih terarah dan mudah untuk dipahami, seperti pada gambar 3.1 dibawa dapat dilihat tahapan penelitian dalam bentuk flowchart yang mana merupakan table dari tahapan penelitian :



**Gambar 3. 1 Flowchart**

### 3.3 Studi Literatur

Dalam penelitian ini, penulis melakukan studi literatur untuk penelitian sejenis yaitu penelitian dengan data mining. Hal ini bertujuan untuk mencari data penguat teori-teori yang ada dalam penelitian ini. Penulis juga memanfaatkan studi literatur untuk mencari berbagai macam referensi dari metode apa saja yang cocok digunakan untuk penelitian ini.

### 3.4 Pengumpulan Data

Dalam melakukan penelitian ini tentunya penulis membutuhkan data-data yang nantinya penulis gunakan sebagai bahan untuk melakukan penelitian. Sesuai tempat penulis melakukan penelitian, , penulis juga mencari data dari sumber terpercaya yaitu Tata Usaha Jurusan Teknologi Informasi. Yang mana data yang didapatkan merupakan data utuh berbentuk *Microsoft excel* sehingga memudahkan penulis dalam melakukan penelitian dimana terdapat 16.907 baris data dan 15 atribut.

**Tabel 3. 1 data awal yang dikumpulkan**

Nomor Mahasiswa	Nama Mahasiswa	Kode Matakuliah	Nama Matakuliah	Kelas	Nilai	No	No Mhs	Nama Mhs	IPK	In Pembimbing	Pen Pembimbing	Judul TA	Nilai	Dosen Penguji 1
		TI1101	Basic English	A	A	1			null	null	null	null	null	null
		TI1102	Agama Islam Agribisnis	A	A	2			null	null	null	null	null	null
		TI1103	Dasar Teknologi Informasi	A	BC	3			null	null	null	null	null	null
		TI1104	Aplikasi Produktivitas	A	A	4			3.41				A	
		TI1104P	Praktikum Aplikasi Produktivitas	A		5			3.53				#REF!	#REF!
		TI1105	Logika dan Teknik Pemrograman	A	D	6			3.24				#REF!	#REF!
		TI1106	Teknik Komputer	A	AB	7			3.57				#REF!	#REF!
		TI1106P	Praktikum Teknik Komputer	A		8			null	null	null	null	null	null
		TI1107	Sistem Operasi	A	A	9			3.52				#REF!	#REF!
		TI1107P	Praktikum Sistem Operasi	A		10			3.48				A	
		TI1108	Perancangan Basis Data	A	A	11			3.22				A	
		TI1101	Basic English	A	A	12			3.82				A	
		TI1102	Agama Islam Agribisnis	A	A	13			3.74				A	
		TI1103	Dasar Teknologi Informasi	A	AB	14			3.86				#REF!	
		TI1104	Aplikasi Produktivitas	A	A	15			null	null	null	null	null	null
		TI1104P	Praktikum Aplikasi Produktivitas	A		16			3.43				A	
		TI1105	Logika dan Teknik Pemrograman	A	D	17			3.53				#REF!	#REF!
		TI1106	Teknik Komputer	A	AB	18			3.56				#REF!	#REF!
		TI1106P	Praktikum Teknik Komputer	A		19			3.25				#REF!	#REF!
		TI1107	Sistem Operasi	A	A	20			null	null	null	null	null	null
		TI1107P	Praktikum Sistem Operasi	A		21			3.65				AB	
		TI1108	Perancangan Basis Data	A	B	22			3.47				#REF!	#REF!

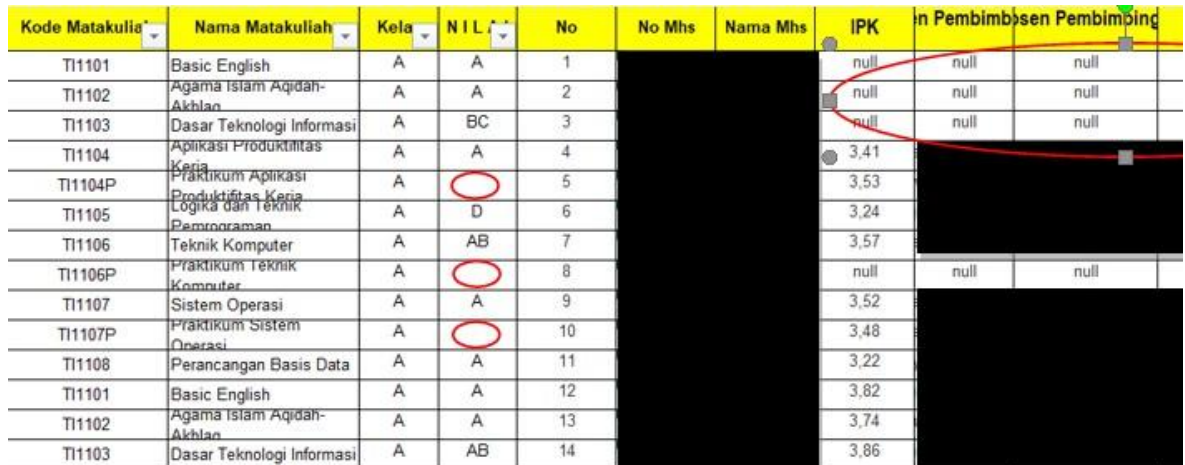
### 3.5 Pre-Processing Data (Pengolahan Data)

Setelah selesai mengumpulkan data pastinya kita akan memiliki sebuah database untuk bahan uji yang akan diproses nanti, tetapi semua data yang sudah ada di database belum tentu semuanya layak atau bisa digunakan. Oleh karena itu diperlukan *pre-processing data*. Pengolahan data dipakai untuk mengekstrak dan membersihkan data dari tidak konsistennya dan tidak kompatibelnya dengan format yang sesuai (Zai, 2022). Proses pengolahan data terbagi menjadi sebagai berikut:

#### 3.5.1 Data Cleaning

proses ini berguna dalam menentukan data mana yang akan digunakan dalam pengolahan data nantinya, dimana penulis memeriksa satu persatu apakah ada data yang nilainya *null* atau data yang *value* tidak sesuai. Beberapa data tabel yang tidak terpakai akan dihilangkan didalam proses

pembersihan data atau *data cleaning*. Seperti pada gambar 3.2 dapat dilihat pada lingkaran merah ada beberapa data di tiap atribut memiliki nilai yang tidak terpakai sehingga dibutuhkan *data cleaning* yang berguna untuk menghapus atau menghilangkan data yang tidak terpakai. Untuk lebih jelasnya dapat dilihat pada gambar di bawah ini:



Kode Matakuliah	Nama Matakuliah	Kelas	Nilai	No	No Mhs	Nama Mhs	IPK	In Pembimbing	Sen Pembimbing
TI1101	Basic English	A	A	1			null	null	null
TI1102	Agama Islam Aqidah-Akhlak	A	A	2			null	null	null
TI1103	Dasar Teknologi Informasi	A	BC	3			null	null	null
TI1104	Aplikasi Produktifitas Kerja	A	A	4			3,41		
TI1104P	Praktikum Aplikasi Produktifitas Kerja	A		5			3,53		
TI1105	Logika dan Teknik Pemrograman	A	D	6			3,24		
TI1106	Teknik Komputer	A	AB	7			3,57		
TI1106P	Praktikum Teknik Komputer	A		8			null	null	null
TI1107	Sistem Operasi	A	A	9			3,52		
TI1107P	Praktikum Sistem Operasi	A		10			3,48		
TI1108	Perancangan Basis Data	A	A	11			3,22		
TI1101	Basic English	A	A	12			3,82		
TI1102	Agama Islam Aqidah-Akhlak	A	A	13			3,74		
TI1103	Dasar Teknologi Informasi	A	AB	14			3,86		

**Gambar 3. 2 data cleaning**

Terlihat masih banyak data yang tidak memiliki nilai atau *null* dan tidak memiliki *value* yang sesuai dengan kasus penulis. pembersihan data sangat penting karena database tidak dirancang khusus untuk analisis bibliometrik. yang mana secara khusus, anda harus menghapus duplikat dan entri yang salah (Naveen Donthu, 2021). Setelah dibersihkan total data adalah 6.262 baris dan 15 atribut, setelah itu penulis dapat melanjutkan ke tahap selanjutnya yaitu *data selection*.

### 3.5.2 Data Selection

*Data Selection* atau bisa dibilang seleksi data adalah tahapan data yang akan melewati tahap seleksi dari atribut-atribut yang tidak terpakai, yang mana atribut yang seleksi tidak berkaitan dengan kasus yang akan diuji oleh peneliti, hasil dari proses *data selection* dapat dilihat pada tabel 3.2 dimana akan terlihat data dan atribut mana yang dipakai dan tidak terpakai.

**Tabel 3.2 data selection**

Atribut yang terpakai	Atribut yang tidak terpakai
Nomor Mahasiswa	Kode Matakuliah
Nama Matakuliah	Kelas

Nilai	No
NIM	Nama Mahasiswa
Judul TA	IPK
	Dosen Pembimbing 1
	Dosen Pembimbing 2
	Dosen Penguji 1

Dapat dilihat terdapat beberapa atribut yang tidak relevan dengan kasus yang sedang diteliti oleh penulis, sehingga harus diseleksi agar mempermudah dalam melakukan pengolahan yang mana pada tahap ini data menyusut ke 6.262 baris dan 5 atribut.

### 3.5.3 Transformasi Data

Tranformasi data atau *Data Transformation* adalah tahapan di mana data ditransformasikan dan dikonsolidasikan ke dalam bentuk yang sesuai untuk mining. Transformasi data diperlukan untuk memudahkan proses klasifikasi menggunakan RapidMiner. Transformasi pada *big data* baru-baru ini telah menghadirkan beberapa tantangan dan peluang substansial untuk pemilihan fitur yang akan digunakan dalam pengolahan data (Jundong Li, 2017). Pada tahapan ini penulis melakukan transformasi terhadap atribut matakuliah yang dijabarkan menjadi 28 atribut, dimana hal itu dapat dilihat pada tabel 3.3 di bawah ini.

***Tabel 3. 3 transformasi atribut matakuliah***

Atribut Awal	Atribut Setelah Transformasi
Matakuliah	Logika dan Teknik Pemrograman
	Sistem Operasi
	Perancangan Basis Data
	Pemrograman Berorientasi Obyek
	Pengembangan Aplikasi Windows
	Implementasi Basis Data
	Probabilitas dan Statistik
	Pemrograman Java
	Analisa dan Perancangan Berorientasi Obyek

	Pengembangan Aplikasi Terintegrasi
	Algoritma dan Struktur Data
	Jaringan Komputer
	Interaksi Manusia dan Komputer
	Pengembangan Konten Web
	Pengembangan Aplikasi Mobile
	Pengembangan Aplikasi Basis Data
	Protokol Routing pada Jaringan Komputer
	Perancangan Sistem Multimedia
	Pengembangan Aplikasi Terdistribusi
	Pengembangan Aplikasi Web
	Pengembangan Web Service
	Jaringan Switching dan Wireless
	Pengembangan Aplikasi Mobile
	Pengembangan Aplikasi Basis Data
	Pengembangan Komponen Web
	Pengujian dan Pejaminan Kualitas Software
	Pengembangan Aplikasi Enterprise
	Teknologi Server
	Jaringan Area Luas

Selanjutnya penulis menambahkan atribut baru yang mana fungsi dari atribut ini adalah sebagai label dari seluruh atribut didatabase, fungsi label itu sendiri adalah sebagai atribut target dari database itu sendiri, untuk lebih jelasnya dapat dilihat di tabel 3.4 di bawah ini.

***Tabel 3. 4 Atribut Label***

<b>Atribut label</b>	<b>Penejelasan</b>
Jaringan	Atribut ini berkaitan dengan judul TA yang berkaitan dengan jaringan
UI/UX	Atribut ini berkaitan dengan judul TA yang berkaitan dengan UI/UX

Data	Atribut ini berkaitan dengan judul TA yang berkaitan dengan Data
Pemrograman	Atribut ini berkaitan dengan judul TA yang berkaitan dengan Pemrograman

Sehingga data yang terkumpul adalah sebanyak 202 baris dan 32 atribut yang akan diolah.

### 3.6 Implementasi Algoritma

Tahapan yang selanjutnya adalah implementasi algoritma, pada tahapan ini penulis melakukan pemodelan data dengan metode yang akan dipakai yaitu *Support Vector Machine*. Data yang telah selesai melewati beberapa tahapan tadi seperti *Data Selection*, *Cleaning* data dan *Transformasi* data akan menghasilkan data yang siap uji untuk nantinya akan diuji menggunakan algoritma *Support Vector Machine* didalam *RapidMiner*. Metode *Support Vector Machine* akan menggunakan *Hyperplan* untuk melakukan pemisahan antara dua kelas dengan jenis yang berbeda sehingga menghasilkan keputusan yang terbaik.

Tentunya sebelum memasukkan data ke dalam *RapidMiner* terlebih dahulu penulis memisahkan data yang tidak terpakai dari hasil *Preprocessing* tadi. Pada pengolahan ini semua atribut data digunakan sebagai data latih dan data uji kemudian penulis melakukan pengujian menggunakan metode *Cross Validation*. Untuk pengujian ini penulis menggunakan operator *Cross Validation*. Pada tahapan ini *Cross Validation* akan melakukan pemisahan pada data latih secara otomatis dan untuk jumlahnya sendiri berapa data training dan berapa data testing yang diinginkan itu tergantung pada seberapa banyak penulis ingin melakukan jumlah pengujian.

Pada pengujian menggunakan *Cross Validation* sendiri terdapat dua parameter yang harus diperhatikan, yaitu :

#### a. *Number of Folds*

parameter ini berguna untuk menentukan seberapa banyak pengulangan

pengujian yang harus dilakukan, seperti contoh jika *number of fold* dimasukkan dengan angka 10 maka jumlah pengulangan pengujian akan dilakukan sebanyak 10 kali. Dan untuk pembagiannya sendiri data mana yang menjadi data testing dan data training dilakukan secara default oleh *Cross Validation* sendiri, dengan komposisi 90% data akan menjadi data training dan 10% data akan menjadi data testing dan *Cross Validation* ini akan terus melakukan pengulangan sebanyak 10 kali sesuai dengan jumlah *Number of Folds* yang telah dimasukkan tadi tetapi di setiap pengulangan akan dilakukan pertukaran data testing dengan data training.

***b. Sampling Type***

Adalah parameter yang berguna untuk menentukan jenis pengambilan data yang akan diolah. Ada beberapa jenis pengambilan data pada *cross validation*, dimana tiap jenis memiliki pola pengambilan data yang berbeda. Jenis pengambilan datanya yaitu:

***a. Shuffled Sampling***

adalah proses pengambilan data secara acak dari data set.

***b. Linear Sampling***

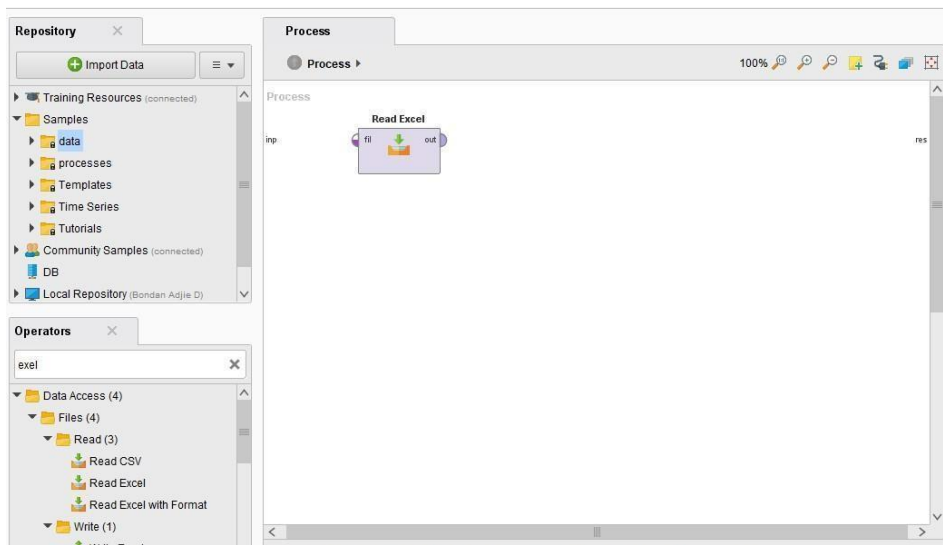
adalah cara pengambilan data secara linear tanpa merubah urutan yang ada pada data set

***c. Stratified Sampling***

adalah proses yang cara pengambilan datanya hampir sama dengan *Shuffled Sampling* bedanya adalah pada *Stratified Sampling* memastikan seluruh distribusi atribut pada subset sama dengan keseluruhan data set.

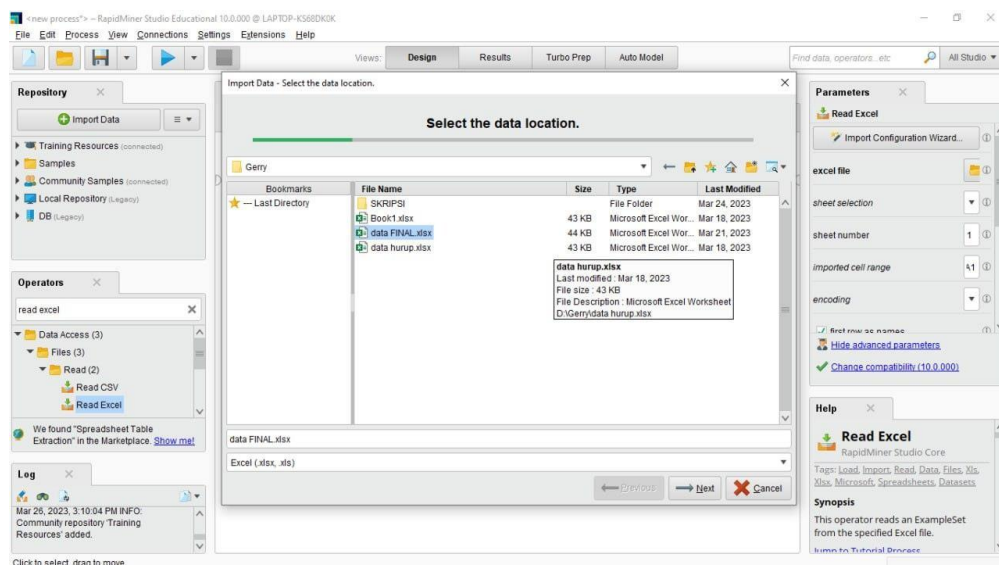


Selanjutnya adalah proses pengujian pada *RapidMiner* mulai dari memasukkan data uji hingga mendapatkan hasil yang diinginkan dari RapidMiner. Langkah pertama yang harus dilakukan yaitu penulis tunjukkan pada Gambar 4.2.



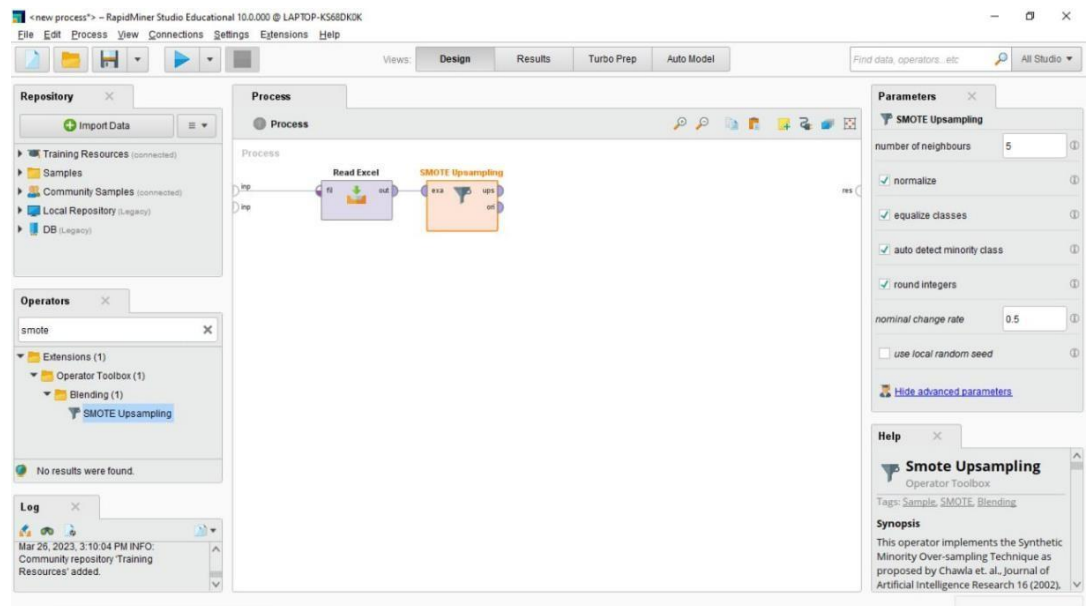
**Gambar 4.1 Proses Drag and Drop Read Excel**

*tool read excel* memiliki fungsi untuk memanggil data yang berbentuk excel yang mana data ini berkesinambungan dengan data yang penulis miliki yang telah melalui berbagai tahapan pengolahan, lalu penulis akan memasukan data excel yang dimiliki dengan cara mengklik *tool read excel* sehingga akan muncul tampilan seperti gambar 4.3 dimana penulis akan memilih data mana yang akan digunakan, untuk lebih jelasnya dapat dilihat digambar di bawah ini.



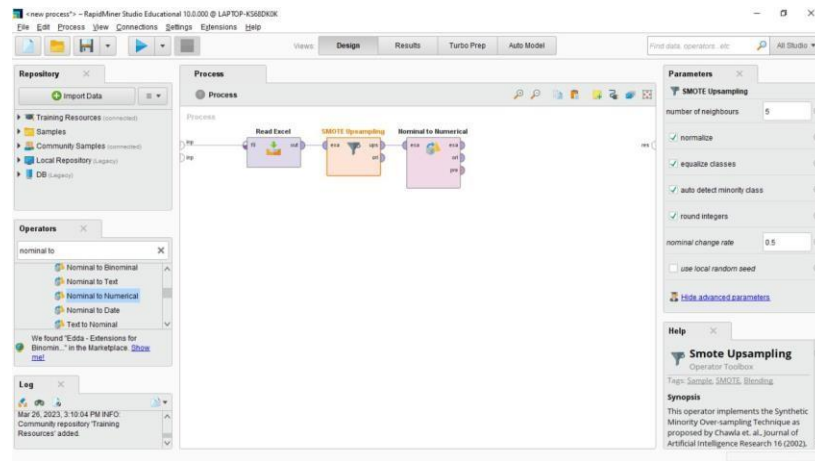
**Gambar 4. 1 Proses Import Excel**

Setelah melakukan *import file excel* langkah selanjutnya adalah menambahkan operator *SOMOTE Upsampling*, operator ini berfungsi untuk mengatasi data yang memiliki value kecil sehingga tidak akan menimbulkan data yang inbalance, SMOTE merupakan salah satu turunan teknik over-sampling untuk menanggulangi ketidakseimbangan kelas dengan menyeimbangkan dataset dengan meningkatkan ukuran kelas minor (Femi Dwi Astuti, 2021) yang mana hal ini dapat dilihat digambar 4.4 di bawah ini.



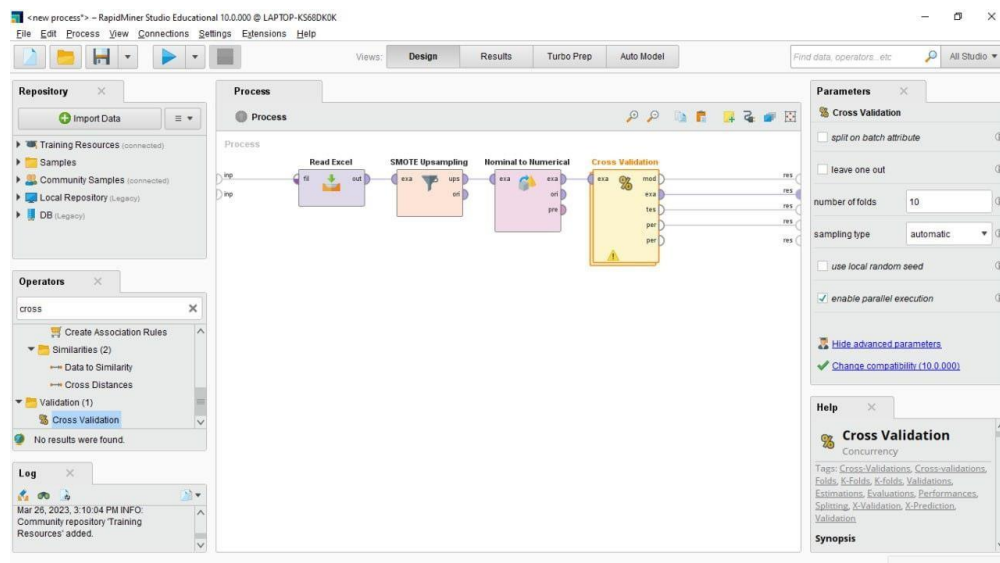
**Gambar 4. 2 Operator SMOTE Upsampling**

Alasan penulis menggunakan Operator SMOTE dikarenakan terdapat label JARINGAN yang hanya memiliki 4 data yang mana, hal ini berbanding jauh dengan label DATA yang memiliki 49 data, label UI/UX yang memiliki 33 data dan label PEMROGRAMAN yang memiliki 116 data, dimana hal ini menyebabkan inbalance data/ Tahap selanjutnya adalah menambahkan operator *Nominal to Numerical* yang mana fungsi dari operator ini adalah melakukan konversi pada tiap-tiap atribut yang bernilai *non-numeric* menjadi *numeric*. Hal ini dapat dilihat digambar 4.5 dibawah ini.



**Gambar 4. 3 Menambahkan Nominal to Numerical**

Langkah selanjutnya adalah menambahkan *operator Cross Validation* yang mana berfungsi untuk menilai kinerja proses sebuah metode algoritme dengan membagi sampel data secara acak dan mengelompokkan data tersebut sebanyak nilai K k-fold, lalu hubungkan tiap-tiap port sesuai dengan Gambar 4.6 di bawah ini.

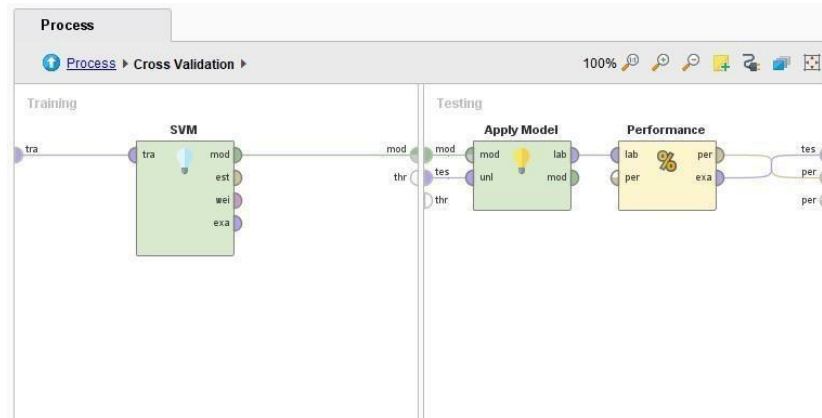


**Gambar 4. 4 Menambahkan Operator Cross Validation**

Setelah *Cross Validation* berhasil ditambahkan selanjutnya adalah melakukan seting dengan cara *double click* pada operator *cross validation*. Setelah masuk ke tab *setting cross validation* langkah selanjutnya adalah menambahkan metode *machine learning* yang diinginkan, dan pada penelitian ini penulis menggunakan metode *Support Vectro Machine*. Tambahkan *support vector machine* pada kolom training, jika kesulitan mencari cukup ketikkan saja nama

operator yang ingin dicari pada tab pencarian *Operators* yang berada diposisi sebelah kiri bawah.

Kemudian untuk dikolom testing lakukan penambahan operator dengan cara yang sama tetapi untuk kolom testing hanya membutuhkan duaoperator yaitu *Apply*



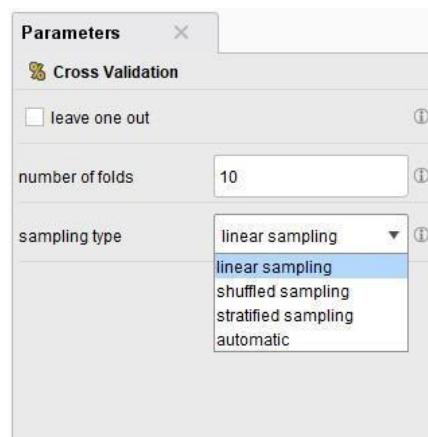
**Gambar 4.5 melakukan setting cross validation**

*Model* dan *Performance*. Setelah semua penambahan operator selesai dilakukan langkah selanjutnya adalah hubungkan tiap-tiap port yang dibutuhkan, bagaimana bentuk penambahan, Kemudian untuk dikolom testing lakukan penambahan operator dengan cara yang sama tetapi untuk kolom testing hanya membutuhkan dua operator yaitu *Apply Model* dan *Performance*. Setelah semua penambahan operator selesai dilakukan langkah selanjutnya adalah hubungkan tiap-tiap port yang dibutuhkan, bagaimana bentuk penambahan operator dan penghubungan port nya akan penulis tunjukan pada Gambar 4.7.

Setelah setting *cross validation* selesai dilakukan langkahselanjutnya adalah melakukan pengaturan parameter *cross validation*. Pada tahapan ini penulis melakukan setting pada *Number of Folds* dimana penulis memasukkan angka 10 yang artinya saat pengujian nanti akan dilakukan pengulangan sebanyak 10 kali. Dan pada parameter *sampling type* penulis menggunakan *linear sampling* terlebih dahulu. Ditahap pengujian yang selanjutnya penulis menggunakan setting yang sama hanyasaja terjadi perubahan pada *sampling type* nya, setelah linear penulis mencoba merubah dengan *shuffled sampling* dan *stratified sampling*.

Contoh pengaturan *number of folds* dan *sampling type* nya akan penulis

tunjukan pada Gambar 4.8.



**Gambar 4. 6 pengaturan parameter**

Setelah semua setting selesai dilakukan pastikan kembali bahwa semua port yang dibutuhkan sudah terhubung dan semua setting telah benar, jika sudah yakin langkah selanjutnya adalah melakukan pengujian dengan cara melakukan klik tombol run yang ada pada *toolbar* yang berada dibagian atas dan biasanya disimbolkan dengan gambar PLAY seperti pada gambar 4.9.



**Gambar 4. 7 run rapidminer**

Tunggu beberapa saat setelah rapidminer selesai melakukan proses uji secara otomatis tabs akan berpindah ke bagian *Result* yang menampilkan hasil dari uji dari proses yang sudah dilakukan menggunakan *Support Vector Machine*.

### 3.7 Analisis Hasil

Dari hasil yang penulis dapatkan dari penelitian ini maka penulis bisa menarik kesimpulan yang mengacu pada rumusan masalah yang sudah dibuat sebelumnya, dan juga dapat membantu saran untuk kemajuan penelitian selanjutnya.

## BAB IV

### HASIL DAN PEMBAHASAN

#### 4.1 Implementasi Algoritma

Pada tahap *performance vector* akan menunjukkan seberapa besar tingkat akurasi menggunakan *support vector machine* (SVM) dengan menggunakan 4 label dapat dilihat pada tabel di bawah ini.

**Tabel 4. 1 Hasil Rapidminer dengan 1 operator SMOTE Upsampling**

	True Pemograman	True Data	True UI/UX	True Jaringan	Class precision
Pred. pemrgmn	114	46	32	0	59.38%
Pred. data	0	0	0	0	0.00%
Pred. UI/UX	0	0	0	0	0.00%
Pred. Jringn	2	3	1	116	95.08%
Class recall	98.28%	0.00%	0.00%	100.00%	

Diatas adalah nilai dengan yang dihasilkan oleh pengolahan data dengan 1 *SMOTE Upsampling* dengan akurasi yang dihasilkan sebesar 73.60%, hasil ini telah diproses oleh sistem *RapidMiner* dengan data yang telah dikumpulkan. Lalu selanjutnya merupakan nilai yang dihasilkan oleh pengolahan data dengan menambahkan 3 Operator *SMOTE Upsampling* dengan masing-masing tiap operator menggunakan 3 *number of neighbours*, dengan *setting* seperti itu maka akan menghasilkan akurasi sebesar 90.74% yang mana merupakan hasil dengan akurasi tertinggi. Untuk lebih jelasnya dapat dilihat pada tabel 4.2 dibawah ini.

**Tabel 4. 2 Hasil Rapidminer dengan 3 operator SMOTE Upsampling**

	True Pemograman	True Data	True UI/UX	True Jaringan	Class precision
Pred. pemrgmn	107	15	14	0	78.68%
Pred. data	6	98	2	0	92.45%
Pred. UI/UX	3	3	100	0	94.34%
Pred. Jringn	0	0	0	116	100.00%

Class recall	92.24%	84.48%	86.21%	100.00%	
--------------	--------	--------	--------	---------	--

Pada bagian *Precision* merupakan class yang mengukur ketepatan informasi dari sistem rapidminer. Sedangkan recall adalah class yang mengukur tingkat keberhasilan proses rapidminer untuk menemukan kembali akurasi. Untuk mencari nilai dari akurasi, precision dan recall dapat dirumuskan dengan rumus seperti pada tabel 4.3 di bawah ini.

**Tabel 4. 3 Rumus Perhitungan Precision, Akurasi dan Recall**

	Nilai Sebenarnya		
		TRUE A	TRUE B
NILAI PREDIKSI	PRED A	AP (A Positive) <i>Corect Result</i>	BP (B Positive) <i>Unexpected result</i>
	PRED B	BN (B Negative) <i>Missing result</i>	AN (A Negaive) <i>Corect obsence result</i>

$$Precision = \frac{AP}{AP + BP}$$

$$Recal = \frac{AP}{AP + BN}$$

$$Accuracy = \frac{AP + AN}{AP + AN + BP + BN}$$

Dari bentuk rumus di atas penulis dapat melakukan perhitungan matematis untuk menghitung akurasi, *precision* dan *recall* terhadap hasil akurasi tertinggi, yang akan penulis tunjukkan seperti berikut :

- a. *Class precision dan recall* yang tepat

*Precision*

$$\frac{107}{107 + 15 + 14 + 0} = 0,78 = 78\%$$

*Recall*

$$\frac{107}{107 + 6 + 3 + 0} = 0,92 = 92\%$$

b. Tingkat akurasi

$$\frac{107 + 98 + 100 + 116}{107 + 98 + 100 + 116 + 15 + 14 + 0 + 6 + 2 + 0 + 3 + 3 + 0 + 0 + 0} \times 100 = 90\%$$

Berdasarkan hasil perhitungan manual di atas penulis berhasil mendapatkan hasil yang sama dengan tabel 4.2, ini membuktikan bahwa perhitungan yang sebelumnya dilakukan oleh RapidMiner ini benar, juga sekaligus membuktikan bahwa pengambilan keputusan menggunakan *Support Vector Machine (SVM)* memiliki tingkat keakurasian yang baik dengan mendapat 90% tingkat akurasi, untuk hasil ketepatan juga memiliki hasil yang baik dalam penggunaan metode ini dan dapat di buktikan dengan *presentase Precision* dan *Recall* yang memiliki angka % tinggi

Lalu pada Tabel 4.1 dapat dilihat bahwa terdapat 2 label yang salah dalam menentukan prediksi, maka dari itu penulis akan mencoba memecahkan masalah di atas dengan melihat apakah prediksi yang dihasilkan dipengaruhi oleh nilai yang diperoleh selama perkuliahan lebih tinggi terhadap label yang tidak sesuai dengan label yang diprediksi berdasarkan judul tugas akhir atau dikarenakan matakuliah yang berkaitan dengan label lebih sedikit dari label yang diprediksi. Sebelum menjelaskan pengaruh nilai matakuliah terhadap prediksi label yang salah disini penulis akan menjelaskan beberapa hasil yang diperoleh menggunakan 2 sampling menggunakan 1 operator SMOTE dan 3 operator SMOTE dan tidak menggunakan SMOTE, sehingga penulis dapat melihat bahwa terdapat perbedaan hasil akurasi dari dua sampling yang sudah diolah, hal ini dapat di ditabel 4.2 di bawah ini.

**Tabel 4. 4 hasil pengujian sampling dan smote**

<b><i>Sampling type</i></b>	<b>Hasil</b>
<i>Linear Sampling (SMOTE) 3</i>	93.52%
<i>Linear Sampling (SMOTE)</i>	73,20%
<i>Linear Sampling</i>	57,43%
<i>Shuffled Sampling (SMOTE) 3</i>	93.74%
<i>Shuffled Sampling (SMOTE)</i>	73.62%



<i>Shuffled Sampling</i>	57,52%
--------------------------	--------

Pada label diatas bahwa perbedaan sampling memiliki pengaruh hasil yang tidak terlalu banyak, hal ini dikarenakan terdapat data yang inbalance sehingga mengurangi hasil akurasi, namu ketika menggunakan opertator 3 SMOTE maupun operator 1 SMOTE terdapat perbedaan hasil akurasi yang sangat signifikan, bahkan perbedaan yang dihasilkan dengan menggunakan 3 operator SMOTE dan tanpa SMOTE sebesar 36.31%, sehingga pilihan menggunakan SMOTE dan SVM adalah pilihan yang tepat.

Selanjutnya penulis akan menjelaskan faktor yang menyebabkan kesalahan prediksi pada label yang terjadi diproses mining, yang dapat dilihat di tabel 4.5 dan 4.6, dimana penulis mengambil satu contoh dari data yang memiliki kesalahan prediksi , yang mana dapat dilihat pada tabel di bawah ini.

***Tabel 4. 5 Kesalahan Prediksi Label Data***

	A	AB	B	BC	C	D	E
DATA	0	1	2	0	1	1	0
PEMROGRAMAN	1	0	3	2	1	5	0

***Tabel 4. 6 Tabel Kesalahan Prediksi Label UI/UX***

	A	AB	B	BC	C	D	E
UI/UX	1	1	2	0	0	0	0
PEMROGRAMAN	1	0	5	0	4	2	0

Dapat dilihat pada tabel diatas, jika melihat dari segi nilai seharusnya tidak akan ada kesalahan prediksi dalam menentukan label hal ini dikarenakan nilai dari 2 label diatas terbilang seimbang, bahkan pada label UI/UX terlihat bahwa matakuliah yang berkaitan UI/UX memiliki nilai yang jauh lebih baik dari pada PEMROGRAMAN. Namun tetap saja terjadi diprediksi sebagai label PEMROGRAMAN, namun jika dilihat Kembali berdasarkan jumlah nilai atau jumlah matakuliah yang berkaitan dengan label diatas, dapat dilihat bahwa label PEMOGRAMAN memiliki jumlah matakuliah yang berkaitan lebih banyak, hal ini dapat menjadi jawaban mengapa terjadi kesalahan prediksi terhadap label di atas, hal ini dikarenakan tidak seimbangnya matakuliah yang berkaitan dengan

label yang diprediksi, sehingga menyebabkan kesalahan prediksi dari prediksi yang sebenarnya.

#### **4.2 Analisis Proses dan Hasil Pengujian**

Didalam proses pada permodelan *data mining*, *data training* dan *data testing* dibagi dan ditentukan secara otomatis oleh operator *cross validation*. *Cross validation* memecah data menjadi data set dengan ukuran yang sama. Penggunaan *SMOTE* terhadap metode *SVM* dilakukan guna mengatasi data yang inbalance, dimana data yang inbalance dapat mengurangi akurasi hasil dari *data testing*. Sedangkan data yang salah prediksi terjadi dikarenakan ketidak seimbangan terhadap jumlah matakuliah yang berkaitan dengan label yang digunakan.

Berdasarkan pada hasil penelitian yang sudah didapatkan, bisa diamati bahwa data yang tidak menggunakan *SMOTE* memiliki akurasi yang jauh lebih kecil dan juga diikuti dengan prediksi, *class recall*, dan *class precision* yang tidak seimbang, hal ini disebabkan oleh data yang inbalance dan juga faktor matakuliah yang berkaitan dengan label yang tidak seimbang. Sehingga dibutuhkan *SMOTE* untuk meningkatkan akurasi dan menyeimbangkan data.

Dan dari penelitian yang penulis lakukan, dapat disimpulkan bahwa penggunaan *SVM* dengan operator *SMOTE Upsampling* yang disertakan *Shuffled Sampling* adalah pengaturan yang memiliki akurasi paling tinggi dibandingkan dengan pengaturan lain.

## **BAB V**

### **KESIMPULAN DAN SARAN**

#### **5.1 Kesimpulan**

Dari hasil pembahasan yang penulis telah lakukan terhadap penelitian ini, maka penulis dapat menyimpulkan sebagai berikut :

1. Metode *support vector machine* dapat digunakan untuk merekomendasi Tugas Akhir Mahasiswa berdasarkan Nilai Matakuliah dan Tugas akhir yang diambil apakah sudah tepat atau tidak.
2. *SMOTE Upsampling* memiliki pengaruh yang signifikan dalam membantu mengatasi data yang *inbalance* dan juga meningkatkan akurasi yang dihasilkan, hal itu terlihat dengan akurasi tertinggi diangka 90.74% yang mana menggunakan *SMOTE Upsampling*.
3. *Sampling type Shuffled* berhasil mendapatkan akurasi tertinggi dengan 90.74% menggunakan *SMOTE Upsampling*. Hal ini dapat ditingkatkan lagi jika memiliki data yang lebih banyak

#### **5.2 Saran**

Saran yang dapat penulis sampaikan dari penelitian ini untuk penelitian selanjutnya yang memiliki kemiripan dengan penelitian yang dilakukan penulis adalah sebagai berikut.

1. Penulis berharap pada penelitian selanjutnya peneliti memiliki lebih banyak faktor yang berpengaruh terhadap kasus yang sedang diemban.
2. Penulis berharap pada penelitian selanjutnya peneliti memiliki lebih banyak data yang dapat diolah, sehingga dapat mendapatkan akurasi yang lebih baik.
3. Penulis berharap pada penelitian selanjutnya peneliti dapat mencari faktor yang mempengaruhi tugas akhir mahasiswa yang didapatkan dari luar lingkungan perkuliahan mahasiswa itu sendiri.
4. Penulis berharap pada penelitian selanjutnya peneliti dapat membuat perbandingan metode SVM dengan metode lainya.

## DAFTAR PUSTAKA

- Auliya Rahman Isnain, A. I. (2021). SENTIMEN ANALISIS PUBLIK TERHADAP KEBIJAKAN LOCKDOWN PEMERINTAH JAKARTA MENGGUNAKAN ALGORITMA SVM.
- Debby Alita, Y. F. (2020). IMPLEMENTASI ALGORITMA MULTICLASS SVM PADA OPINI PUBLIK BERBAHASA INDONESIA DI TWITTER.
- Fathurrohlim, Q. (2020). Porting LIBSVM ke node sensor Preon32.
- Femi Dwi Astuti, F. N. (2021). Implementasi SMOTE untuk mengatasi ImbalanceClass pada Klasifikasi Car Evolution menggunakan K-NN.
- Fitri Pratiwi, P. Y. (2019). Implementasi Data Mining dalam Pemilihan Bahasa Pemrograman dalam Penyelesaian Tugas Akhir Mahasiswa. *Seminar Nasional Industri dan Teknologi (SNIT), Politeknik Negeri Bengkalis*.
- fiyanti, S. D. (2015). PENENTUAN PEMINATAN TUGAS AKHIR MAHASISWA TEKNIK INFORMATIKA UNSIKA.
- Jumeilah, F. S. (2017). Penerapan Support Vector Machine (SVM) untuk Pengkategorian Penelitian.
- Jundong Li, K. C. (2017). Feature Selection: A Data Perspective.
- Muhammad Zamroni Uska, R. H. (2020). Evaluation of Rapidminer-Aplication in Data Mining Learning using PeRSIVA Model.
- Nameirakpam Dhanachandra, Y. J. (2020). An image segmentation approach based on fuzzy c-means and dynamic particle swarm optimization algorithm.
- Naveen Donthu, S. K. (2021). How to conduct a bibliometric analysis: An overview and guidelines.
- Nurul Renaningtias, D. A. (2021). PENERAPAN METODE PROTOTYPE PADA PENGEMBANGAN SISTEM INFORMASI TUGAS AKHIR MAHASISWA.
- Sagala, R. M. (2021). Prediksi Kelulusan Mahasiswa Menggunakan Data mining Algoritma K-means.
- Utpal Barman, R. D. (2020). Soil texture classification using multi class support vector machine.
- Zai, C. (2022). IMPLEMENTASI DATA MINING SEBAGAI PENGOLAHAN DATA.