UNIVERSITAS
MUHAMMADIYAH
SIDOARJO

UMSIDA
DARI SINI PENCERAHAN BERSEMI

Penggunaan Datamining Untuk Memprediksi Masa Studi Mahasiswa di Universitas Muhammadiyah Sidoarjo Dengan Algoritma Naive Bayes

Ujian Skripsi

Nama :Muhammad Mursidil Arif
Prodi :Informatika
Nim :191080200034
Angkatan : 2019

Dosen Pembimbing : Hamzah Setiawan, S.Kom., M.Kom
Dosen Penguji 1 : Arif Senja Fitrani, M. Kom
Dosen Penguji 2 : Yunianita Rahmawati, S.Kom., M.Kom

PENDAHULUAN

Datamining merupakan salah satu metode yang sering digunakan untuk mendapatkan informasi penting bagi penggunanya. Nama lain dari datamining adalah knowledge discovery in databases (KDD), yang artinya adalah proses pengumpulan data dan penggunaan data masa lampau, untuk mendapatkan informasi dari sebuah database.

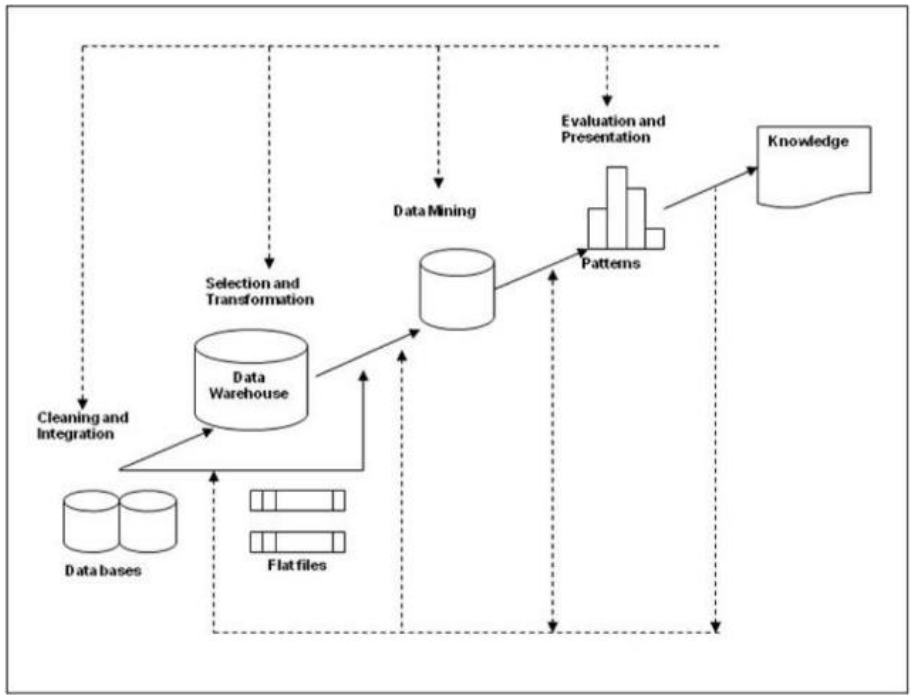
Fungsi datamining secara umum adalah classification, clustering, Associasion, Regression, Forecasting, Sequence Analysis, dan Deviation Analysis. cara kerja klasifikasi terbagi menjadi dua tahap penting, yaitu :

1. Pembuatan model yang nantinya disimpan sebagai prototype memori.

2. Menggunakan model yang disimpan untuk melakukan proses klasifikasi pada objek data lain dan menetapkan data tersebut sesuai kelasnya .

METODE

Tahapan penelitian dilakukan dengan menjalankan langkah-langkah dari proses knowledge discovery in database (KDD), langkah-langkah tersebut yaitu : pengumpulan data, preprocessing data, permodelan naïve bayes, testing model, dan evaluasi.



Hasil dan Pembahasan

Pengumpulan Data

Proses pengumpulan data pada penelitian ini dilakukan dengan menggunakan data alumni program studi informatika universitas Muhammadiyah sidoarjo angkatan 2017-2018 yang didapatkan dari Direktorat Sistem dan Teknologi Informasi (DSTI) Universitas Muhammadiyah Sidoarjo sebanyak 200 data dengan 9 atribut numerik dan 1 label output. Pada gambar 3 berikut adalah tampilan dataset yang akan digunakan.

	ips1	ips2	ips3	ips4	ipk	sekolah asal	keuangan	beasiswa	kemahasiswaan	kelulusan
0	3.38	3.41	3.81	3.71	3.5775	1	0	1	1	1
1	3.42	3.29	3.58	3.47	3.4400	1	0	1	1	0
2	3.32	3.30	3.31	3.36	3.3225	0	0	1	1	1
3	3.32	3.26	3.17	3.25	3.2500	0	0	1	1	1
4	3.65	3.38	3.64	3.64	3.5775	1	0	1	1	1
5	3.43	3.22	3.60	3.35	3.4000	1	0	1	1	0
6	3.37	3.30	3.57	3.34	3.3950	1	0	1	1	1
7	3.47	3.28	3.47	3.23	3.3625	1	0	1	1	1
8	3.47	3.51	3.68	3.49	3.5375	1	0	1	0	0
9	3.37	3.35	3.60	3.32	3.4100	1	0	1	1	1

Proses Permodelan Naïve Bayes dengan Python

Proses klasifikasi naïve bayes dilakukan menggunakan aplikasi anaconda dengan tools jupyter notebook dan python sebagai bahasa pemrogramannya. yang pertama kali harus dilakukan adalah menyiapkan data dari Microsoft excel dengan format .xlsx, setelah itu mengimport library python yang diperlukan seperti *(pandas, numpy, matplotlib)*.

Hasil dan pembahasan

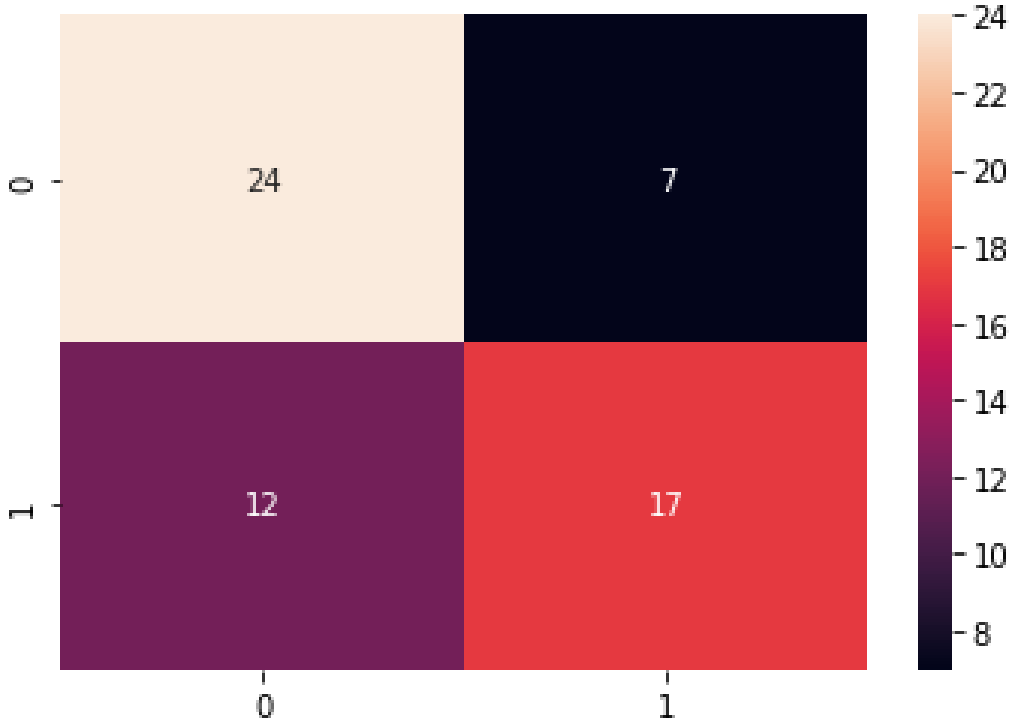
Perancangan sistem

- **Konsep**
- Dalam menentukan konsep Aplikasi Pengenalan Virus Covid 19 Berbasis Augmented Reality Sebagai Media pengenalan virus covid 19 Berbasis *Augmented Reality*
- **Tabel 1. Konsep aplikasi**

Judul	Tujuan	Pengguna	Objek	Masukan	Keluaran
Aplikasi Pengenalan Virus Covid 19 Berbasis Augmented Reality	Membuat aplikasi yang dapat mempermudah pengenalan virus covid 19	Publik	Virus covid19 varian Omikron varian Alpa varian Delta	Gambar	Audio,dan animasi objek tiga dimensi Virus covid 19 varian Omikron varian Alpa varian Delta

Hasil Pengujian Confusion Matrix

Setelah dilakukan proses permodelan naïve bayes dengan bahasa pemrograman python pada tools jupyter notebook, maka didapatkan hasil sebagai berikut :



Hasil Pengujian Confusion Matrix

• Setelah mendapatkan nilai dari confusion matrix, pengujian model naïve bayes pada penelitian ini mendapatkan hasil accuracy 68%, dengan precision 0.67, recall 0.77 dan F1-score 0.72. Sedangkan accuracy score mendapatkan nilai 68.33%.

	precision	recall	f1-score	support
Tepat	0.67	0.77	0.72	31
Tidak Tepat	0.71	0.59	0.64	29
accuracy			0.68	60
macro avg	0.69	0.68	0.68	60
weighted avg	0.69	0.68	0.68	60

SIMPULAN

• Pada penelitian penggunaan datamining untuk memprediksi masa studi mahasiswa universitas Muhammadiyah sidoarjo dengan algoritma naïve bayes ini, didapatkan model dengan nilai accuracy score sebesar 68.33%. Sedangkan evaluasi menggunakan confusion matrix didapatkan hasil accuracy 68% dengan precision 0.67, recall 0.77 dan f1-score 0.72 untuk kelas tepat waktu, dan precision 0.71, recall 0.59, f1-score 0.64 untuk kelas tidak tepat waktu. Penelitian ini menggunakan data dari alumni program studi informatika angkatan 2017-2018 sebanyak 200 data dengan 9 variabel prediktor dan 1 variabel target.

UMSIDA
DARI SINI PENCERAHAN BERSEMI