

BAB II

TINJAUAN PUSTAKA

II.1. Penelitian Terkait

Telah ada beberapa penelitian yang terkait dengan judul data mining yang menggunakan metode algoritma K-Nearest Neighbor (KNN), diantaranya adalah :

1. Penelitian yang dilakukan Novita Marina (2015) dengan judul penerapan algoritma K-NN (*nearest Neighbor*) untuk deteksi penyakit (Kanker Serviks). Hasil dari penelitian ini adalah Penerapan Algoritma k-NN (*nearest Neighbor*) Untuk Deteksi Penyakit (Kanker Serviks). Sistem pakar merupakan sistem yang berusaha mengadopsi kepakaran manusia sehingga komputer bisa melakukan hal-hal yang dapat dikerjakan oleh seorang pakar untuk memecahkan permasalahan yang bersifat spesifik. Pakar dalam hal ini adalah seorang yang ahli dibidangnya. Sistem pakar dapat digunakan untuk semua bidang ilmu termasuk dunia medis/kedokteran. Salah satu yang berkaitan dengan medis adalah penyakit kanker mulut rahim yang amat ditakutkan semua wanita karena menyerang organ reproduksi yang disebabkan oleh virus *human virus papilloma* (HPV).
2. Penelitian yang dilakukan oleh Ashar Jihar (2016) dengan judul penelitian implementasi metode K-Nearest Neighbor (K-NN) dan Simple Additive Weighting (SAW) dalam pengambilan keputusan seleksi penerimaan anggota paskibra. Hasil dari penelitian ini adalah Implementasi Metode K-Nearest Neighbor (K-NN) Dan Simple Additive Weighting (SAW) Dalam

Pengambilan Keputusan Seleksi Penerimaan Anggota Paskibra. Penelitian ini membangun sebuah sistem pendukung keputusan seleksi penerimaan anggota Paskibra. Aplikasi yang dibangun menggunakan metode k-Nearest Neighbor (KNN) dan Simple Additive Weighting (SAW). Metode k-Nearest Neighbor digunakan untuk melakukan klasifikasi peserta yang akan diterima. Metode Simple Additive Weighting digunakan untuk melakukan perankingan.

3. Penelitian yang dilakukan oleh Jodi Irjaya Kartika (2017) dengan judul penelitian penentuan siswa berprestasi menggunakan metode K-Nearest Neighbor dan Weighted Product (Studi Kasus : SMP Negeri 3 Mejayan). Hasil dari penelitian ini adalah Penentuan Siswa Berprestasi Menggunakan Metode K-Nearest Neighbor dan Weighted Product (Studi Kasus : SMP Negeri 3 Mejayan). Pendidikan mempunyai peranan yang sangat penting untuk kemajuan bangsa ini, Sekolah sebagai institusi pendidikan, mengembangkan berbagai sistem pembinaan yang sifatnya memotivasi dan mengembangkan potensi siswa. Salah satunya dengan melakukan pemilihan siswa berprestasi. Namun pada proses menentukan siswa berprestasi hanya dinilai berdasarkan aspek akademik saja.
4. Sedangkan penelitian yang dilakukan oleh penulis meneliti disektor pemerintahan, pengguna sistem informasi sangatlah penting dalam rangka mendukung kegiatan pekerjaan yang ada disektor pemerintahan. Dalam mendukung kegiatan pekerjaan yang ada disektor pemerintahan diperlukan manajemen yang baik. Sejalan dengan perkembangan pengetahuan dan

kemajuan teknologi yang semakin pesat, sehingga menuntun kita untuk membuka wawasan diri dan menerima perubahan-perubahan yang terjadi di negara kita, baik pengaruh dari dalam maupun dari luar negara kita sendiri. Sedangkan pada Dinas Pendidikan PROVSU khususnya di bidang hukum masih menerapkan sistem manual yaitu dengan melakukan pencocokan data dengan data pegawai yang melakukan pelanggaran, pihak manajemen sektor pemerinthan harus mencocokkan hukuman yang pas pada pegawai yang melakukan pelanggaran disiplin dan dalam pengolahan data pelanggaran disiplin yang dilakukan oleh pegawai yaitu masih menerapkan sistem *microsoft excel* sehingga memakan banyak waktu dan tenaga, hal ini dinilai kurang efektif dan tidak efisien serta berisiko terjadinya kerusakan pada data. Sistem informasi penentuan sangatlah diperlukan oleh suatu instansi disektor pemerintahan. Dalam penentuan pemberian sanksi disiplin pegawai biasanya akan terdapat beberapa kandidat dengan nilai yang berbeda pada setiap kriteria, dimana masing-masing kriteria memiliki prioritas yang berbeda-beda. Hal tersebut seringkali menjadi permasalahan yang cukup kompleks. Hasil akhir yang diperoleh dari perhitungan adalah nilai total dari setiap kandidat untuk seluruh kriteria.

II.2. Data Mining

Data Mining merupakan disiplin ilmu yang mempelajari metode untuk mengekstrak pengetahuan atau menemukan pola dari suatu data. Data Mining adalah suatu metode pengolahan data untuk menemukan pola yang tersembunyi

dari data tersebut. Hasil dari pengolahan data dengan metode data mining ini dapat digunakan untuk mengambil keputusan di masa depan. Data mining ini juga dikenal dengan istilah *pattern recognition*.

Menyebutkan bahwa KDD atau *Knowledge Discovery From Data*, merupakan proses terstruktur, yaitu sebagai berikut:

1. *Data Cleaning* adalah proses memberikan data dari data *noise* dan tidak konsisten.
2. *Data Integration* adalah proses untuk menggabungkan data dari beberapa sumber yang berbeda.
3. *Data Selection* adalah proses untuk memilih data dari *database* yang sesuai dengan tujuan analisis.
4. *Data Transformation* adalah proses mengubah bentuk data menjadi data yang sesuai untuk proses mining.
5. *Data Mining* adalah proses penting yang menggunakan sebuah metode tertentu untuk memperoleh sebuah pola dari data.
6. *Pattern Evaluation* adalah proses mengidentifikasi pola.
7. *Knowledge Presentation* adalah yang dapat merepresentasikan informasi yang dibutuhkan, proses dimana informasi yang telah didapatkan kemudian digunakan oleh pemilik data (Heni Sulastri;2017:2).

II.4. Metode K-NN (K-Nearest Neighbor)

K-Nearest Neighbor (K-NN) adalah suatu metode yang menggunakan algoritma supervised dimana hasil dari query instance yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada K-NN. Tujuan dari algoritma ini adalah mengklasifikasikan obyek baru berdasarkan atribut dan training sample. Classifier tidak menggunakan model apapun untuk dicocokkan dan hanya berdasarkan pada memori. Diberikan titik query akan ditemukan sejumlah obyek atau (titik training) yang paling dekat dengan titik query. Klasifikasi menggunakan voting terbanyak diantara klasifikasi dari k obyek. Algoritma K-NN menggunakan klasifikasi ketetangga sebagai nilai prediksi dari query instance yang baru. Algoritma metode K-NN sangatlah sederhana, bekerja berdasarkan jarak terpendek dari query instance ke training sample untuk menentukan K-NN-nya. Training sample diproyeksikan ke ruang berdimensi banyak, dimana masing-masing dimensi merepresentasikan fitur dari data. Ruang ini dibagi menjadi bagian-bagian berdasarkan klasifikasi training sample. Sebuah titik pada ruang ini ditandai kelas c jika kelas c merupakan klasifikasi yang paling banyak ditemui pada k buah tetangga terdekat dari titik tersebut. Dekat atau jauhnya tetangga biasanya dihitung berdasarkan Euclidean Distance yang direpresentasikan pada persamaan 1 sebagai berikut :

$$D(a, b) = \sqrt{\sum_{k=1}^d (a_k - b_k)^2}$$

Dimana matriks $D(a,b)$ adalah jarak skalar dari kedua vector a dan b dari matriks dengan ukuran dimensi, (Jodi Irjaya; 2017 : 354).

II.4.1. Conhto Kasus Metode K-NN (K-Nearest Neighbor)

Penelitian yang dilakukan Rhman Rosyida (2019) dengan judul penelitian Perbandingan Algoritma K-Nn Dan Cart Pada Data Mining Penerimaan Beasiswa, pada penelitian ini penulis menggunakan aplikasi WEKA untuk melakukan perbandingan data, adapun penjelasannya dapat dilihat sebagai berikut :

1. Pengumpulan Data

Pengumpulan Data Sumber data utama adalah bagian kemahasiswaan STMIK AMIKOM Purwokerto yang menyediakan data penerima beasiswa. Terdapat 8 atribut / fitur dalam data tersebut, yang terdiri dari 7 atribut kriteria dan 1 atribut keputusan. Adapun atribut-atribut yang digunakan adalah sebagai berikut : jenis kelamin, semester, IPK, pekerjaan orang tua, jumlah anggota keluarga, penghasilan orang tua, prestasi dan status. Data ini harus diolah terlebih dahulu melalui tahap pre-procesing, dimana tahapan ini untuk menyesuaikan atribut-atribut yang akan digunakan dalam mengolah dataset tersebut. Berikut ini adalah tabel 1 dataset pendaftar beasiswa STMIK AMIKOM Purwokerto yang belum dilakukan penyesuaian.

Tabel.II.1. Dataset Sebelum Penyesuaian

Jenis Kelamin	Semester	IPK	Pekerjaan Orang Tua	Jumlah Anggota Keluarga	Penghasilan (Rp)	Prestasi	Status
P	4	3.82	Wiraswasta	3	1.300.000	-	Ditolak
P	4	3.53	Wiraswasta	4	1.500.000	-	Ditolak
L	6	3.51	Wiraswasta	4	2.000.000	-	Ditolak
P	6	3.66	Petani	4	1.500.000	-	Ditolak
L	6	3.60	Swasta	7	650.000	-	Diterima
L	6	3.46	Buruh Tani	4	300.000	-	Diterima

L	6	3.36	Pedagang	4	200.000	-	Diterima
L	2	3.25	Pedagang	6	1.500.000	-	Ditolak
L	6	3.73	PNS	4	1.500.000	-	Ditolak
P	4	3.66	Buruh	5	1.000.000	-	Ditolak
P	6	3.10	Pedagang	4	1.500.000	-	Ditolak
P	6	3.28	Wiraswasta	4	1.750.000	-	Ditolak
L	6	3.51	Buruh	4	1.500.000	-	Ditolak
L	6	3.72	Supir	3	1.000.000	-	Diterima
L	4	3.59	PNS	4	303.6000	-	Ditolak
P	2	3.38	Buruh	3	1.000.000	-	Diterima
P	2	3.38	Polri	4	2.220.000	-	Ditolak
P	2	3.38	Buruh Tani	9	-	-	Ditolak
L	2	3.46	Wiraswasta	8	1.500.000	-	Ditolak
L	2	3.75	TNI	7	3.762.2228	-	Ditolak
...
...
L	6	2.92	Swasta	4	2.500.000	-	Ditolak
L	6	3.48	BUMN	5	3.000.000	-	Ditolak
P	6	3.57	Swasta	4	1.300.000	-	Ditolak
L	2	3.21	Penjahit	-	2.500.000	-	Ditolak

2. Tahap Pre-Processing

Awal dataset terdiri dari 150 data dengan 8 atribut yaitu jenis kelamin, semester, IPK, pekerjaan orang tua, jumlah anggota keluarga, penghasilan, prestasi, dan status. Jumlah mahasiswa yang diterima 44 dan yang ditolak berjumlah 106. Terdapat data yang tidak lengkap yaitu 1 data berasal dari atribut jumlah anggota keluarga dan 7 data dari atribut penghasilan orang tua. Atribut jenis kelamin, semester, IPK, pekerjaan orang tua dan status memiliki nilai yang lengkap. Diasumsikan pekerjaan orang tua dibedakan menjadi dua yaitu bekerja dan tidak bekerja. Untuk prestasi, data yang kosong diasumsikan tidak mempunyai prestasi. Setelah melakukan penanganan missing value, data menjadi 142 dengan jumlah kasus yang diterima

berjumlah 41, dan yang ditolak berjumlah 101. Proses dikritisasi akan dilakukan untuk mempermudah pengelompokan nilai dan mempersempit permasalahan serta meningkatkan keakurasian (Lesmana, 2012). Berikut ini adalah penyesuaian atribut yang digunakan untuk mengolah data pada tabel 2.

Tabel II.2. Atribut/Fitur Penghasilan

Atribute	Keterangan	Nilai	Nilai Baru
Penghasilan Orang Tua	Berisikan besarnya penghasilan orang tua mahasiswa	$\leq 1.500.000$	Rendah
		1.500.000 – 2.500.000	Sedang
		2.500.000 – 3.500.000	Tinggi
		$\geq 3.500.000$	Sangat Tinggi

Setelah dilakukan proses pre-processing, data berjumlah 142 dengan 41 mahasiswa yang diterima dan 101 yang ditolak. Berikut adalah dataset yang siap digunakan dalam aplikasi Weka pada tabel 3.

Tabel II.3. Hasil Preprocessing Data

Jenis Kelamin	Semester	IPK	Pekerjaan Orang Tua	Jumlah Anggota Keluarga	Penghasilan (Rp)	Prestasi	Status
P	4	3.82	Bekerja	3	Rendah	Tidak	Ditolak
P	4	3.53	Bekerja	4	Rendah	Tidak	Ditolak
L	6	3.51	Bekerja	4	Sedang	Tidak	Ditolak
P	6	3.66	Bekerja	4	Rendah	Tidak	Ditolak
L	6	3.60	Bekerja	7	Rendah	Ya	Diterima
L	6	3.46	Bekerja	4	Rendah	Ya	Diterima
L	6	3.36	Bekerja	4	Rendah	Tidak	Diterima
L	2	3.25	Bekerja	6	Rendah	Tidak	Ditolak
L	6	3.73	Bekerja	4	Rendah	Ya	Ditolak
P	4	3.66	Bekerja	5	Rendah	Ya	Ditolak
P	6	3.10	Bekerja	4	Rendah	Tidak	Ditolak
P	6	3.28	Bekerja	4	Sedang	Tidak	Ditolak

L	6	3.51	Bekerja	4	Rendah	Ya	Ditolak
L	6	3.72	Bekerja	3	Rendah	Tidak	Diterima
L	4	3.59	Bekerja	4	Sangat Tinggi	Tidak	Ditolak
P	2	3.38	Bekerja	3	Sedang	Tidak	Diterima
P	2	3.38	Bekerja	4	Sangat Tinggi	Tidak	Ditolak
P	2	3.38	Bekerja	9	Sangat Tinggi	Tidak	Ditolak
L	2	3.46	Bekerja	8	Tinggi	Tidak	Ditolak
L	2	3.75	Bekerja	7	Rendah	Tidak	Ditolak
...
...
L	6	2.92	Bekerja	4	Sedang	Tidak	Ditolak
L	6	3.48	Bekerja	5	Tinggi	Tidak	Ditolak
P	6	3.57	Bekerja	4	Rendah	Ya	Ditolak
L	2	3.21	Bekerja	-	Rendah	Ya	Ditolak

3. Penggunaan Metode Klasifikasi

Setelah tahap pre-processing selesai kemudian dataset tersebut mulai diolah dengan aplikasi Weka. Tahapan ini juga bertujuan menghasilkan confusion matrix dan melakukan 2 kali percobaan, percobaan pertama dengan metode evaluasi 10-fold cross validation , dimana dataset dibagi menjadi 10 subsets (9 subsets sebagai training sets dan 1 subsets sebagai testing sets) dengan jumlah 10 kali iterasi, dan yang kedua dengan use training set. Metode K-NN akan dicoba dengan 142 dataset dengan 6 kali percobaan sehingga nilai k yang digunakan dari 1 sampai 6. Percobaan pertama menggunakan 10-fold cross validation dan percobaan kedua menggunakan use training set. Hasil akurasi ditunjukkan pada tabel 4 berikut ini :

Tabel III.4. Perbandingan Akurasi

Hasil Akurasi		
Nilai K	Menggunakan 10-fold cross validation	Menggunakan use training set
K1	64.7887%	99.2958%

K2	46.4789%	83.8028%
K3	57.0423%	80.2817%
K4	54.2254%	72.5352%
K5	64.7887%	75.3521%
K6	59.8592%	75.3521%

Nilai akurasi berdasarkan uji coba dataset pendaftar beasiswa sangat dipengaruhi nilai k. Nilai k yang tinggi berakibat pada semakin banyak tetangga dalam proses klasifikasi dan noise semakin tinggi. Diketahui hasil akurasi yang terbaik pada nilai k-1 yaitu sebesar 99.2958 % dengan menggunakan use training set. Berikut ditunjukkan hasil output clasifier pada weka secara rinci. Perhitungan hasil akurasi berdasarkan precision, recall, dan F-measure tercantum dalam tabel berikut :

Tabel II.5. Of Confusion Kelas Diterma

41 (True Positive)	0 (False Negative)
1 (False Positive)	100 (True Negative)

Dalam persamaa (1)

$$Precision = \frac{TP}{TP + FP} = \frac{41}{41 + 1} = 0.976$$

$$Recall = \frac{TP}{TP + FN} = \frac{41}{41 + 0} = 1$$

$$F - measure = \frac{2 \times precision \times recall}{precision + recall} = \frac{2 \times 0.976 \times 1}{0.976 + 1} = 0.988$$

Kelas “Ditolak”

Berikut adalah tabel 6 yang menggambarkan of confusion kelas “ditolak”:

Tabel II.6. Of Confusion Kelas Ditolak

100 (True Positive)	1 (False Negative)
0 (False Positive)	41 (True Negative)

Dalam persamaan (2)

$$Precision = \frac{TP}{TP + FP} = \frac{100}{100 + 0} = 1$$

$$Recall = \frac{TP}{TP + FN} = \frac{100}{100 + 1} = 0.99$$

$$F - measure = \frac{2 \times precision \times recall}{precision + recall} = \frac{2 \times 1 \times 0.99}{1 + 0.99} = 0.995$$

Dari hasil precision, recall, dan F-measure kelas Diterima dan Ditolak, dapat dihitung nilai rata-rata dari kelas kelas-kelas yang ada (Weighted Avg) dengan terlebih dulu menjumlahkan nilai A = (41 + 0) = 41 dan B = (100 + 1) = 101. Rumusnya sebagai berikut :

Dalam persamaan (3)

$$Weighted\ Avg\ (precision) = \frac{0.976 \times 41 + 1 \times 101}{142} = 0.993$$

$$Weighted\ Avg\ (recall) = \frac{1 \times 41 + 0.99 \times 101}{142} = 0.993$$

$$Weighted\ Avg\ (F - measure) = \frac{0.988 \times 41 + 0.995 \times 101}{142} = 0.993$$

Nilai akurasi confusion matrix dapat dilihat pada tabel 7.

Tabel II.7. Nilai Akurasi Berdasarkan Confusion Matrix

Class	Precision	Recall	F-Measure
Diterima	0.976	1	0.988
Ditolak	1	0.99	0.995
Weighted Avg	0.993	0.993	0.993

II.5. Pelanggaran Kode Etik

Pelanggaran Kode Etik, adalah pelanggaran yang dilakukan oleh anggota terhadap ketentuan yang diatur dalam kode etik tersebut. Sejumlah pelanggaran kode etik yang dilanggar oleh Anggota DPR RI adalah Pasal 2 Peraturan DPR Nomor 01 Tahun 2015 Tentang Kode Etik, (Nur Qamariah Novitas; 2016: 144), yang menyatakan bahwa :

1. Anggota dalam setiap tindakannya harus mengutamakan kepentingan bangsa dan negara daripada kepentingan pribadi, seseorang, dan golongan.
2. Anggota bertanggung jawab mengemban amanat rakyat, melaksanakan tugasnya secara adil, mematuhi hukum, menghormati keberadaan lembaga legislatif, dan mempergunakan fungsi, tugas, dan wewenang yang diberikan kepadanya demi kepentingan dan kesejahteraan rakyat.
3. Anggota harus selalu menjaga harkat, martabat, kehormatan, citra, dan kredibilitas dalam melaksanakan fungsi, tugas, dan wewenangnya serta dalam menjalankan kebebasannya menggunakan hak berekspresi, beragama, berserikat, berkumpul, dan mengeluarkan pikiran dengan lisan dan tulisan.
4. Anggota yang ikut serta dalam kegiatan organisasi di luar DPR, harus mengutamakan tugasnya sebagai Anggota. Pasal 3 Bagian Kedua Tentang Integritas, yang menyatakan bahwa ayat 1 :
 - a. Anggota harus menghindari perilaku tidak pantas, atau tidak patut yang dapat merendahkan citra dan kehormatan DPR, baik di dalam gedung

DPR maupun di luar gedung DPR, menurut pandangan etika dan norma yang berlaku dalam masyarakat.

- b. Anggota sebagai wakil rakyat, memiliki pembatasan pribadi dalam bersikap, bertindak, dan berperilaku.
- c. Anggota harus menjaga nama baik dan kewibawaan DPR.
- d. Anggota dilarang meminta dan menerima pemberian atau hadiah selain dari apa yang berhak diterimanya sesuai dengan ketentuan peraturan perundang-undangan.

Pasal 4 Bagian Ketiga Tentang Hubungan dengan Mitra Kerja menyatakan bahwa ayat :

1. Anggota harus bersikap profesional dalam melakukan hubungan dengan Mitra Kerja.
2. Anggota dilarang melakukan hubungan dengan Mitra Kerjanya untuk maksud tertentu yang mengandung potensi korupsi, kolusi dan nepotisme. Pasal 6 Bagian Kelima, Tentang Keterbukaan dan Konflik Kepentingan menyatakan bahwa ayat : 4 Anggota dilarang menggunakan jabatannya untuk mencari kemudahan dan keuntungan pribadi, Keluarga, Sanak Famili, dan golongan. (Lihat Peraturan DPR Nomor 01 Tahun 2015 Tentang Kode Etik DPR RI). , (Nur Qamariah Novitas; 2016: 144).

II.6. Sistem Pembuktian Pelanggaran Kode Etik

Sistem Pembuktian Pelanggaran Kode Etik Sistem merupakan kumpulan dari beberapa bagian yang memiliki keterkaitan dan saling bekerja sama serta

membentuk suatu kesatuan untuk mencapai suatu tujuan dari sistem tersebut. Maksud dari suatu sistem adalah untuk mencapai suatu tujuan dan sasarannya. Arti penting pembuktian adalah, mencari kebenaran atas suatu peristiwa hukum. Dalam hukum acara pidana, pembuktian merupakan inti persidangan perkara pidana, karena yang dicari dalam perkara pidana adalah kebenaran materil. Pembuktiannya dimulai sejak tahap penyelidikan, untuk mencari, dan menemukan peristiwa yang diduga sebagai tindak pidana. Pada tahap ini, sudah terjadi pembuktian, dengan tindakan penyidik dalam mencari barang bukti, maksudnya untuk membuat terang/jelas suatu tindak pidana, serta menentukan atau menemukan tersangkanya. Tentunya, dengan cara-cara, serta penilaian alat bukti yang telah ditentukan oleh Undang-Undang. Terlihat begitu rumit dan bertingkat-tingkatnya pembuktian hukum pidana, jika hukum Acara pidana ini, digunakan untuk menyelesaikan pelanggaran kode etik, jelas akan memakan waktu yang lama, sehingga tujuan utamanya untuk menyelamatkan lembaga/ institusi jabatan yang tercoreng oleh oknum-oknum dari kepercayaan publik tidak tercapai. Ada beberapa sistem pembuktian yang dikenal dalam doktrin hukum acara pidana, (Nur Qamariah Novitas; 2016: 146), yaitu :

1. Teori pembuktian yang mana hakim terikat pada alat bukti berdasarkan Undang-Undang. (Positief wettelijk bewijstheorie) Artinya jika dalam pertimbangan, hakim telah menganggap suatu perbuatan sesuai dengan alat-alat bukti yang ditentukan dalam Undang-Undang tanpa diperlukan keyakinan hakim, Positief wettelijk bewijstheorie ini digunakan dalam hukum acara perdata.

2. Teori pembuktian berdasarkan keyakinan hakim semata, (Conviction-in time), teori ini menitik beratkan pada hati nurani hakim. Terbukti atau tidaknya ke salahannya terdakwa, semata-mata ditentukan atas penilaian keyakinan atau perasaan hakim.
3. Teori pembuktian berdasar keyakinan hakim dalam batas-batas tertentu, atas alasan-alasan yang logis (Conviction raisonee).
4. Teori pembuktian Negatief wettelijk bewijstheorie, berdasarkan pembuktian menurut keyakinan hakim yang timbul dari alat-alat bukti yang ditentukan dalam undang-undang secara negatif. Secara tegas dasar pembuktian ini dinyatakan dalam pasal 183 KUHAP, “Hakim tidak boleh menjatuhkan pidana kepada seorang, kecuali apabila dengan sekurang-kurangnya dua alat bukti yang sah. ia memperoleh keyakinan bahwa suatu tindak pidana benar-benar terjadi dan bahwa terdakwa yang bersalah melakukannya”. Teori ini mewajibkan hakim untuk memutuskan perkara dengan menjatuhkan kepada terdakwa dengan berdasarkan pada banyaknya jenis dan jumlah alat bukti yang sah. (Nur Qamariah Novitas; 2016: 146).

II.7. Sistem

Sistem merupakan kumpulan elemen yang saling berhubungan satu sama lain yang membentuk satu kesatuan dalam usaha mencapai satu tujuan. Di dalam perusahaan, yang dimaksud elemen dari sistem adalah departemen-departemen internal seperti persediaan barang mentah, produksi, persediaan barang jadi, promosi, penjualan, keuangan, personalia, serta pihak eksternal seperti *supplier*,

dan konsumen yang saling terkait satu sama lain dan membentuk suatu kesatuan usaha, (Kardiawan Lius Sarumaha; 2014 : 64).

II.8. Sistem Informasi

Sistem Informasi adalah kumpulan elemen yang saling berhubungan satu sama lain yang berbentuk satu kesatuan untuk mengintegrasikan data, memproses dan menyimpan serta mendistribusikan informasi. Sistem informasi dapat didefinisikan sebagai suatu sistem yang dibuat oleh manusia yang terdiri dari beberapa komponen dalam organisasi untuk mencapai suatu tujuan yaitu menyajikan informasi. Komponen sistem informasi terdiri dari :

- a. *Hardware* (perangkat keras), terdiri dari komputer, printer dan jaringan.
- b. *Software*, kumpulan perintah yang ditulis dengan aturan untuk memerintah komputer melaksanakan tugas tertentu.
- c. *Data*, merupakan komponen dasar dari informasi yang akan diproses lebih lanjut untuk menghasilkan informasi.
- d. *Manusia*, yang terlibat dalam komponen manusia seperti operator dan pimpinan.
- e. *Prosedur*, dokumentasi proses sistem, buku penuntun operasional (aplikasi) dan teknis, (Nursahid; 2015: 56).

II.9. Data Dan Informasi

Data merupakan deskripsi tentang benda, kejadian, aktivitas, dan transaksi yang tidak mempunyai makna atau tidak berpengaruh secara langsung kepada makna pemakai. Data juga dapat diartikan suatu bahan mentah yang kelak dapat diolah lebih lanjut untuk menjadi sesuatu yang lebih bermakna. Dan data inilah yang nantinya akan disimpan dalam database. Sedangkan informasi adalah data yang telah diolah menjadi sebuah bentuk yang berarti bagi penerimanya dan bermanfaat dalam pengambilan keputusan saat ini atau saat mendatang, (Muhammad Taufiq; 2013 : 50).

II.10. WEKA (Waikato Environment for Knowledge Analysis)

WEKA (Waikato Environment for Knowledge Analysis) merupakan perangkat lunak data mining yang dikembangkan oleh Universitas Waikato, New Zealand. Diimplementasikan pertama kali pada tahun 1997 dan mulai menjadi open source pada tahun 1999. Hingga saat ini Weka sudah mencapai versi 3.6.11 dengan berbagai pengembangan dari versi pertama 3.3. Ditulis dalam bahasa pemrograman Java, Weka juga didukung oleh GUI yang sangat baik dan user friendly, dapat mengolah berbagai file data seperti *.csv dan *.arff serta memiliki fitur utama seperti data preprocessing tools, learning algorithms dan berbagai metode evaluasi. Selain itu, Weka juga dapat memberikan hasil dalam bentuk visual, seperti tabel dan kurva, Weka terdiri dari beberapa tools yang dapat digunakan untuk melakukan tugas pre-processing data, classification, regression, klustering, association rules, dan visualisasi (Tari Mardiana; 2015: 2).