

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

## BAB II

### LANDASAN TEORI

#### 2.1 Beasiswa

Menurut Murniasih (2009) beasiswa diartikan sebagai bentuk penghargaan yang diberikan kepada individu agar dapat melanjutkan pendidikan ke jenjang yang lebih tinggi. Penghargaan itu dapat berupa akses tertentu pada suatu institusi atau penghargaan berupa bantuan keuangan. Pada dasarnya, beasiswa adalah penghasilan bagi yang menerimanya.

##### 2.1.1. Jenis-Jenis Beasiswa

Menurut Murniasih (2009), ada beberapa jenis beasiswa yaitu:

##### 1. Beasiswa Penghargaan

Beasiswa ini biasanya diberikan kepada kandidat yang memiliki keunggulan akademik. Beasiswa ini diberikan berdasarkan prestasi akademik mereka secara keseluruhan. Misalnya, dalam bentuk Indeks Prestasi Kumulatif (IPK). Meski sangat kompetitif, beasiswa ini ada dalam berbagai bentuk.

##### 2. Beasiswa Bantuan

Jenis beasiswa ini adalah untuk mendanai kegiatan akademik para mahasiswa yang kurang beruntung, tetapi memiliki prestasi. Komite beasiswa biasanya memberikan beberapa penilaian pada kesulitan ini, misalnya, seperti pendapatan orangtua, jumlah saudara kandung yang sama-sama tengah menempuh studi, pengeluaran, biaya hidup, dan lain-lain.

##### 3. Beasiswa Atletik

Universitas biasanya merekrut atlet populer untuk diberikan beasiswa dan dijadikan tim atletik perguruan tinggi mereka. Banyak atlet menyelesaikan pendidikan mereka secara gratis, tetapi membayarnya dengan prestasi olahraga.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Beasiswa seperti ini biasanya tidak perlu dikejar, karena akan diberikan kepada mereka yang memiliki prestasi.

4. Beasiswa Penuh

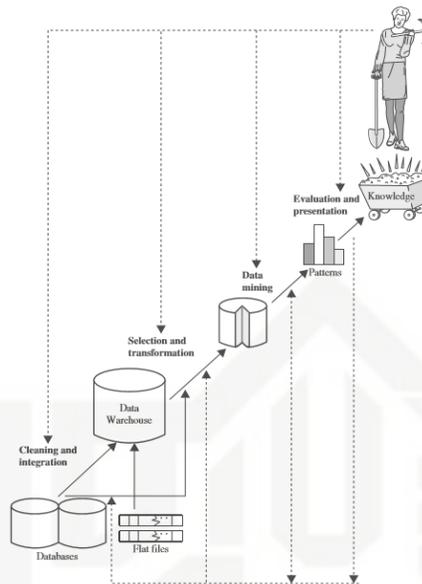
Banyak orang menilai bahwa beasiswa diberikan kepada penerimanya untuk menutupi keperluan akademik secara keseluruhan. Jika Anda benar-benar beruntung, tentunya Anda akan mendapatkan beasiswa seperti ini. Beasiswa akan diberikan untuk menutupi kebutuhan hidup, buku, dan biaya pendidikan. Tetapi, banyak beasiswa lainnya meliputi biaya hidup, buku, atau sebagian dari uang sekolah.

**2.2 Data Mining**

Pengertian lain menurut (Larose, 2005) *data mining* adalah bidang dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, database dan visualisasi untuk penanganan permasalahan pengambilan informasi dari database yang besar. Jadi, *data mining* atau yang disebut dengan *Knowledge Discovery in Data* sering digunakan secara bergantian untuk menjelaskan informasi tersembunyi dalam database besar.

Istilah *Knowledge Discovery in Data* (KDD) dan *data mining* sering digunakan secara bergantian untuk memaparkan proses-proses penggalian informasi yang tersembunyi di dalam kumpulan data dengan basis yang besar atau banyak (Han et al., 2012). KDD dan *data mining* saling berkaitan dimana salah satu tahapan dalam seluruh proses KDD adalah *data mining*. Gambar 2.1 berikut menjelaskan tahapan-tahapan KDD :





**Gambar 2.1 Skema tahapan KDD (Han dkk, 2012)**

Berikut tahapan-tahapan KDD yang terlihat pada gambar 2.1 di atas:

1. Pembersih data (*Data cleaning*)  
 Pembersihan data dilakukan untuk menghilangkan *noise* dan data yang tidak konsisten. Tahap ini dapat dilakukan dengan cara mengisi data yang memiliki *missing value*, memperbaiki kesalahan data, mengidentifikasi dan menghilangkan *outliers*, serta memeriksa data yang tidak konsisten.
2. Integrasi data (*data integration*)  
 Integrasi data merupakan penggabungan beberapa data dari berbagai database kedalam satu database baru. Tidak jarang data yang diperlukan untuk *data mining* tidak hanya berasal dari suatu database tetapi juga berasal dari beberapa database atau file teks.
3. Seleksi data (*data selection*)  
 Tahap ini dilakukan dengan mereduksi data dan mengambil data yang relevan untuk dianalisa.





Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

### 2.2.1 Evaluasi

Evaluasi adalah kunci dalam pembuatan aplikasi berbasis *data mining*. Performa dari suatu model kasifikasi dapat diukur dengan tingkat akurasinya berdasarkan *Confusion matrix*. *Confusion matrix* merupakan alat yang berguna untuk menganalisis seberapa baik *classifier* mengenali tuple dari kelas yang berbeda. TP dan TN memberikan informasi ketika *classifier* benar, sedangkan FP dan FN memberikan informasi ketika *classifier* salah (Han, Kamber, & Pei, 2012). Gambar 2.2 adalah contoh dari *confusion matrix*.

		Predicted class		Total
		yes	no	
Actual class	yes	TP	FN	P
	no	FP	TN	N
Total		P'	N'	P + N

Gambar 2.2 *Confusion Matrix* (Han et al., 2012, p366)

Keterangan :

TP = *true positive*

TN = *true negative*

FP = *false positive*

FN = *false negative*

Dalam *data mining* ada beberapa pola keluaran yang digunakan untuk menguji metode yang digunakan seperti akurasi, *precision*, *recall*, *absolute error*, *clasification error*, dan *root mean squared error*. Akurasi didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai aktual.

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \dots\dots\dots(2.2)$$

*Recall* adalah tingkat keberhasilan sistem dalam sebuah menemukan kembali sebuah informasi.

$$\text{Recall} = \frac{TP}{TP+FN} \dots\dots\dots(2.3)$$



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

*Precision* adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem.

$$Precision = \frac{TP}{TP+FP} \dots\dots\dots(2.4)$$

*Classification error* adalah besar kesalahan yang dihasilkan dari klasifikasi. Semakin kecil nilai APER maka hasil pengklasifikasian semakin besar.

$$APER = \frac{FN+PN}{TP+TN+FP+FN} \dots\dots\dots(2.5)$$

*Absolute error* adalah besar kesalahan antara hasil klasifikasi dengan nilai sebenarnya.

$$AE = f_i - y_i \dots\dots\dots(2.6)$$

Keterangan:

$f_i$  = nilai hasil klasifikasi

$y_i$  = nilai sebenarnya

*Root mean squared error* adalah menentukan seberapa besar kesalahan yang dihasilkan dari model dalam analisa.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - t_i)^2}{n}} \dots\dots\dots(2.7)$$

Keterangan:

$y_i$  = data sebenarnya

$t_i$  = data estimasi

$n$  = jumlah data

### 2.3 Klasifikasi

Sebuah sistem yang melakukan proses klasifikasi diharapkan dapat menentukan semua target data *input* dengan benar, namun tidak dapat dimungkiri bahwa kinerja suatu sistem tidak bisa 100% benar, sehingga sebuah sistem klasifikasi juga harus diukur kinerjanya. Umumnya, pengukuran kinerja klasifikasi dapat dilakukan dengan menggunakan matriks konfusi (*confusion matrix*). Dengan





Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

dibedakan menjadi dua bagian, yaitu variabel respon (*response*) atau biasa juga disebut variabel bergantung (*dependent variable*) serta variabel *explonary* atau bisa juga disebut variabel penduga (*predictor variable*) atau disebut juga variabel bebas (*independent variable*) (Xin Yan dan Xiao Gang Su, 2009).

Regresi linier adalah metode statistika yang digunakan untuk membentuk model hubungan antara variabel terikat (*dependen*) dengan satu atau lebih variabel bebas (*independen*). Alur proses dalam regresi linier (Heri Susanto, 2014) adalah:

1. Menghitung rata-rata setiap atribut

$$\bar{X} = \sum_{i=1}^n x_i \dots\dots\dots(2.8)$$

Keterangan :

- $\bar{X}$ = rata-rata
- n= banyak atribut
- i= nilai

2. Menghitung standar deviasi

$$S = \sqrt{\frac{\sum(x_i - \bar{x})^2}{(n-1)}} \dots\dots\dots(2.9)$$

Keterangan :

- S= standar deviasi
- $x_i$ = nilai x ke i
- $\bar{x}$ = rata-rata
- n= banyak atribut

3. Menghitung nilai a dan  $b_n$

Langkah yang harus dilakukan adalah mencari nilai a dan  $b_n$  agar bisa didapatkan persamaan regresi linier

$$a = \frac{Det(A_0)}{Det(A)} \dots\dots\dots(2.10)$$

$$b_i = \frac{Det(A_i)}{Det(A)} \dots\dots\dots(2.11)$$

Keterangan :

- a = konstanta



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

b = nilai koefisien regresi

Det= determinan

4. Membuat persamaan regresi linier

Setelah didapat nilai a dan  $b_n$  maka lakukan perhitungan dengan menggunakan persamaan garis regresi linier

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_nX_n \dots\dots\dots(2.12)$$

Keterangan :

Y = variable dependen (nilai yang diprediksi)

a = konstanta

b = kemiringan/koefisien korelasi (nilai peningkatan atau penurunan)

x = variable independen

5. Uji korelasi parsial

$$r_{x_{iy}} = \frac{n \sum x_i y - (\sum y \sum x_1)}{\sqrt{[(n \sum y^2) - (\sum y)^2][n \sum x_1^2 - (\sum x_1)^2]}} \dots\dots\dots(2.13)$$

Keterangan :

$r_{x_{iy}}$  = uji korelasi parsial

n= banyak atribut

$\sum$ = jumlah  $x_i y$

6. Koefisien korelasi

$$\sum y^2 = \sum (y^2) - \frac{(\sum y)^2}{n} \dots\dots\dots(2.14)$$

$$\sum x_i y = \sum (x_i y) - \frac{(\sum x_i)(\sum y)}{n} \dots\dots\dots(2.15)$$

Keterangan :

$\sum$ =jumlah

n= banyak atribut

7. Menghitung sumbangan relatif dari kedua prediktor ( $R^2$ )

$$R^2 = \frac{JK_{reg}}{\sum y_i^2} \dots\dots\dots(2.16)$$

Keterangan :

JKreg (jumlah kuadrat regresi)=  $a_1 \sum x_{1i} y_i + a_2 \sum x_{2i} y_i$



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$JK_{res} = \sum (y_i - \bar{y}_i)^2$$

8. Menghitung R simultan

$$R_{simultan} = \sqrt{R^2} \dots \dots \dots (2.17)$$

Keterangan :

$R^2$  = sumbangan relatif dari kedua prediktor

9. Mencari nilai koefisien determinan

$$KP = R_{simultan} \times 100\% \dots \dots \dots (2.18)$$

Kelebihan dari Regresi Linier yaitu dapat mengetahui besarnya pengaruh secara kuantitatif setiap variable bebas apabila pengaruh variable dianggap konstan dan dapat menganalisis dengan menggunakan beberapa variable bebas sehingga hasil yang didapat lebih akurat. Sedangkan kekurangan dari Regresi Linier adalah sebagai berikut:

- a. Data yang di ukur harus linier untuk memperoleh hasil yang baik.
- b. Tidak mampu menunjukkan titik jenuh fungsi yang sedang diselidiki karena selalu timbul kesalahan peramalan.
- c. Terjadinya multikolinieritas pada variabel-variabel yang dipakai dan tidak mampu menjelaskan *variable* tak bebas atau hubungan X dan Y yang tidak bermakna.

**2.5 K-Nearest Neighbor**

*K-Nearest Neighbor* merupakan salah satu metode yang digunakan dalam pengklasifikasian. Prinsip kerja *K-Nearest Neighbor* (KNN) adalah mencari jarak terdekat antara data yang akan dievaluasi dengan *K-Nearest Neighbor* terdekatnya dalam data pelatihan (Agusta, 2007).

Ada banyak cara untuk mengukur jarak kedekatan antara data baru dengan data lama (*data training*), diantaranya *euclidean distance* dan *manhattan distance* (*city block distance*), yang paling sering digunakan adalah *euclidean distance* (Han et



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah,
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

al., 2012). Berikut adalah langkah-langkah dalam algoritma *K-Nearest Neighbor* dengan pencarian jarak menggunakan rumus *Euclidian*:

1. Menentukan Nilai N buah data awal sebagai basis pengetahuan awal sistem.
2. Menentukan Nilai K (tetangga) terdekat.
3. Mempersiapkan data tranning berupa nilai kriteria suatu data baru yang belum diketahui statusnya.
4. Menentukan Status setiap data tranning berdasarkan suatu aturan (*roole*) tertentu untuk menghasilkan basis pengetahuan sistem.
5. Menghitung jarak setiap sampel data tranning terhadap data yang akan diuji (data uji) berdasarkan persamaan:

$$d(x,y) = \sqrt{\sum_i^n (x_i - y_i)^2} \dots\dots\dots(2.19)$$

Keterangan :

d = jarak antara titik pada *data training* x dan titik data testing y yang akan diklasifikasi, dimana  $x = x_1, x_2 \dots, x_i$  dan  $y = y_1, y_2 \dots, y_i$

i = merepresentasikan nilai atribut

n = merupakan dimensi atribut.

6. Menetapkan status data uji berdasarkan nilai rata-rata K buah sampel data tranning terdekat.

Kelebihan *K-Nearest Neighbor* yaitu bahwa dia tangguh terhadap *training* data yang *noise* atau jarak data yang terlalu jauh dan efektif apabila data latihnya besar. Sedangkan kelemahan *K-Nearest Neighbor* adalah :

- a. *K-Nearest Neighbor* perlu menentukan nilai dari parameter k (jumlah dari tetangga terdekat).



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

- b. Pembelajaran berdasarkan jarak tidak jelas mengenai jenis jarak apa yang harus digunakan dan atribut mana yang harus digunakan untuk mendapatkan hasil yang terbaik.
- c. Biaya komputasi cukup tinggi karena diperlukan perhitungan jarak dari tiap sampel uji pada keseluruhan sampel latih.

### 2.6 Rapid Miner

Sebelumnya YALE (*Yet Another Learning Environment*) dikembangkan di *Technical University of Dortmund* pada tahun 2001 oleh Ralf Klinkenberg, Ingo Mierswa dan Simon Fischer. Setelah, nama *software* ini berubah pada tahun 2007 dari YALE ke RapidMiner dan dikembangkan oleh perusahaan RapidMiner, Jerman. RapidMiner adalah sistem berbasis sumber java terbuka untuk pertambangan data dan menyediakan lingkungan yang terintegrasi untuk *machine learning*, *data mining*, *text mining*, analisis prediktif dan analisis bisnis dan terutama digunakan untuk aplikasi bisnis dan industri. RapidMiner adalah *software tool open source* yang paling kuat, mudah digunakan dan intuitive antarmuka *Graphical User* untuk desain proses analitik, yang berisi beberapa "operator" .suatu fungsi operator sebagai satu tugas dalam proses mereka di mana *input* dihasilkan oleh *output* yang ada dari operator. Format data terhubung dengan sangat baik dengan berbagai macam sumber data Oracle, IBM DB2, Microsoft SQL Server, Mysql, Excel, Access, SPSS dan berbagai format lain (Neha Chauhan dan Nisha Gautam, 2015).

### 2.7 Kajian Pustaka

Berikut beberapa penelitian yang berkaitan dengan penelitian ini:

Tabel 2.1 Kajian Pustaka

No	Nama Peneliti	Judul	Metode	Hasil
1.	Fachruddin dan Yovi Pratama	Eksperimen Seleksi Fitur Pada Parameter Proyek Untuk Software Effort Estimation dengan K-Nearest Neighbors"x	KNN	Estimasi terbaik atau nilai RAE dan RMSE terkecil dihasilkan oleh hasil seleksi fitur menggunakan information gain pada dataset china. Dengan nilai RAE 0.0041



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

				dan RMSE 0.0495.
2.	Sefto Pratama, S.Kom, M.Kom	Prediksi Harga Tanah Menggunakan Algoritma Linear Regression	Regresi Linier	Algoritma Linear Regression dengan hasil Root Mean Square Error (RMSE) terendah berada di eksperimen ketiga yang menggunakan data training 70% dan data testing 30% dengan K-Fold validasi 8 dengan hasil 77072562.116 dari atribut label Penawaran M2 yang menjadi acuan harga tanah ideal dikota Banjarmasin
3.	Ihsanul Fikri, S.Kom, K.Kom	“Prediksi Rentet Waktu Pemasaran Printer Canon Menggunakan Neural Network Dan K-Nearest Neighbor	KNN	Algoritma K-NN lebih baik dari Neural Network dengan hasil Root Mean Squared Error (RMSE) untuk NN adalah 92.118 dan K-NN adalah 94.236.
4.	Didik Srianto dan Edy Mulyanto	Perbandingan K-Nearest Neighbor Dan Naive Bayes Untuk Klasifikasi Tanah Layak Tanam Pohon Jati	KNN	Accuracy k-NN lebih tinggi dari pada accuracy Naive Bayes dengan perbandingan 96.66 (k-NN) dan 82.63% (Naive Bayes). Sedangkan pada bagian precision menunjukkan hasil k-NN yang lebih tinggi dari pada Naive Bayes dengan perbandingan 95.45% (k-NN) dan 84.57% (Naive Bayes). Pada bagian recall, k-NN kembali memperoleh hasil yang lebih tinggi dari pada Naive Bayes dengan perbandingan 98.63 (k-NN) dan 82.02% (Naive Bayes).
5.	Ahmad Nouvel	Klasifikasi Kendaraan Roda Empat Berbasis KNN	KNN	Jumlah data sebanyak 14 memiliki tingkat akurasi 78,57% dan RMSE dari 0,23, sedangkan jumlah data sebanyak 1728 telah mencapai 95,78% tingkat akurasi, RMSE 0,19 dan ROC daerah 0.99. Menunjukkan semakin besar jumlah data semakin tinggi accuraynya tingkat.
6.	Arun Sharma et al	<i>An Empirical Study of Supervised Learning Techniques on Multispectral Dataset</i>	KNN	Analisis <i>dataset</i> kompleks memiliki lebih dari 2 kelas dengan 4435 untuk <i>training set</i> dan 2000 untuk pengujian <i>set</i> . Dalam penelitian tersebut menggunakan beberapa teknik yang digunakan



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

				seperti jaringan saraf tiruan, <i>Support Vector Machine</i> , <i>K-Nearest Neighbor</i> , <i>Classifiers Diskriminan</i> dan <i>Naive Bayes</i> . Di antara teknik ini, <i>K-Nearest Neighbor</i> telah memberikan hasil terbaik dengan akurasi 90,3% untuk multispektral <i>dataset</i> citra satelit.
7.	Leidiyana	Penerapan Algoritma <i>K-Nearest Neighbor</i> Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor	KNN	Nilai accuracy 81.46% dan termasuk klasifikasi sangat baik karena memiliki nilai AUC antara 0.90-1.00, yaitu sebesar 0.98.
8.	Bahar dan Nidia Rosmawati	Model Penentu Jenis Beasiswa menggunakan Algoritma K-NN Kombinasi Basis Aturan dan Basis Pengaturan	KNN	Bahwa lebih fleksibel dalam menetapkan status data uji.
9.	Mustafidah	Model Regresi <i>Data Mining</i> Motivasi Belajar Pengaruhnya Terhadap Tingkat Kedisiplinan Mahasiswa	Regresi Linier	Penelitian dalam kasus pengaruh motivasi belajar mahasiswa terhadap tingkat kedisiplinan dalam mengikuti perkuliahan di perguruan tinggi dicoba menggunakan dua macam model regresi yaitu linier dan kuadrat. Hasil yang diperoleh adalah model regresi linier memberikan <i>error</i> (mse) sebesar 8,13% sedangkan model regresi kuadrat memberikan <i>error</i> sebesar 8,15%. Hal ini bisa disimpulkan bahwa regresi linier lebih cocok digunakan dalam kasus tersebut.
10.	Sumarlin	Implementasi Algoritma <i>K-Nearest Neighbor</i> Sebagai Pendukung Keputusan Klasifikasi Penerima Beasiswa PPA dan BBM	KNN	Dari 227 <i>record dataset</i> beasiswa PPA diperoleh tingkat accuracy mencapai 77,96%, sedangkan untuk beasiswa BBM digunakan 183 <i>record dataset</i> dan diperoleh tingkat accuracy mencapai 97,28%. Untuk gabungan beasiswa PPA dan BBM diperoleh accuracy mencapai 85,56%.
11.	Khalid Alkhatib, et el	Stock Price Using KNN Algorith	KNN	Data diambil dari enam perusahaan besar di yordania. Algoritma KNN sangat kuat karena memiliki nilai



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

				error yang kecil. Data yang diambil setiap perusahaan ada 200 data dan pelatihan yang dilakukan menggunakan k 5, adapun root mean squared error sebesar 0.0300.
12.	Firas H. Thanoon	Robust Regression by Least Absolute Deviations Method	Regresi Linier	Pada data 50 memiliki absolute error sebesar 0.764 dan mean squared error sebesar 0.86 dengan distribusi normal.
13.	Hui Li, et el	Integrative Method Based on Linear Regression for the Prediction of Zinc-Binding Sites in Proteins	Regresi Linier	Menggunakan kumpulan data Zhao_dataset, area di bawah kurva presisi ingat (AURPC) Prediktor mencapai hampir 0,9 dan meningkat sebesar 2% -9% dibandingkan dengan tiga prediktor lainnya. Dengan recall 70% dan precision sebesar 94.5%.
14.	Amrin	Data mining dengan Regresi Linier untuk peramalan tingkat inflansi	Regresi Linier	Data yang digunakan adalah data skunder yang diambil dari data inflansi bulanan Indonesia menurut kelompok pengeluaran oleh BPS melalui website resmi, data tersebut sebanyak 118 record dengan data latih 80% dan data uji 20%. Performa model regresi linier berganda dalam meramalkan tingkat inflansi menghasilkan Absolute error sebesar 0.038 dan root mean square error sebesar 0.0481.