# **BAB II LANDASAN TEORI**

## ***State Of The Art***

Berikut ini beberapa penelitian yang telah dilakukan sebelumnya dimana berkaitan dengan topik yang sama namun implementasi algoritma yang berbeda , begitu juga berkaitan dengan algoritma yang sama namun penerapan kasus yang berbeda.

Caesaredi Rama Raharya, Nurul Hidayat, Edy Santoso (2018) dalam jurnalnya yang berjudul “Penentuan Penerimaan Beasiswa Menggunakan Metode *Modified K-Nearest Neighbor”*. Pada penilitian ini menjelaskan masalah mengenai kesulitan dalam seleksi penerima beasiswa, maka dari itu dalam mengatasi masalah tersebut dibuat sistem yang memudahkan petugas dalam seleksi beasiswa dengan menggunakan algoritma *Modified K-nearest Neighbor.* Dari hasil penggunaan algoritma MKNN didapatkan tingkat akurasi sebesar 87.2%[1].

Muhammad Fadhlil Aziz, Sofi Defiyanti, Betha Nurina Sari (2017) dalam jurnalnya yang berjudul “Perbandingan Algoritma Cart dan K-nearest Neighbor Untuk Prediksi Luas Lahan Panen Tanaman Padi di Kabupaten Karawang”. Dalam penelitiannya menggunakan cross validation dalam mennetukan kinerja algoritma manaa yang paling bagus. Hasil dari perbandingan ini menunjukan bahwa algoritma CART bekerja lebih baik dari algortima KNN dengan dengan correlation coefficient 0,9646, MAE 498,6229, dan RMSE 834,0204[2].

Fakihatin Wafiyah, Nurul Hidayat, Rizal Setya Perdana (2017) dalam jurnalnya yang berjudul Implementasi Algoritma Modified K-Nearest Neighbor (MKNN) untuk Klasifikasi Penyakit Demam. Masalah yang terjadi dalam penilitian ini adalah kemiripan gejala demam, sehingga menyulitkan dalam mendapatkan dignosa, dalam menangani masalah tersebut maka digunakan algoritma Modified K-nearest neighbour. Hasil dari penerapan algoritma tersebut rata rata tingkat akurasi 92.42%[5].

Ade Irpan Ramdani (2018) dalam skripsinya yang berjudul “Implementasi Metode Klasifikasi KNN Pada Sistem Rekomendasi Bidang Keahlian Mahasiswa: Studi kasus Teknik Informatika UIN Sunan Gunung Djati Bandung”. Penelitian ini membahas sistem rekomendsi bidang keahlian mahasiswa, dimana memperbaiki penelitian sebelumnya yang dilakukan Muhammad Fikri Lazuardi yang memperoleh hasil 68%. Dalam mengatasinya, maka digunakan penerapan metode KNN. Hasil pengujian pada penelitian ini didapatkan tingkat akurasi 98,30%[6].

Tabel 2.1 State Of The Art

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| No | Peneliti | Judul | Masalah | Metode | Objek Penelitian | Hasil |
| 1. | Caesaredi Rama Raharya, Nurul Hidayat, Edy Santoso | Penentuan Penerimaan Beasiswa Menggunakan Metode *Modified K-Nearest Neighbor* | Kesulitan petugas dalam seleksi penerima beasiswa | *Modified K-Nearest Neighbor* | Beasiswa | Metode mknn menyelesaikan masalah dengan tingkat akurasi 87.2%. |
|  | Muhammad Fadhlil Aziz, Sofi Defiyanti, Betha Nurina Sari | Perbandingan Algoritma Cart dan K-nearest Neighbor Untuk Prediksi Luas Lahan Panen Tanaman Padi di Kabupaten Karawang. | Kesulitan dalam memprediksi luas lahan panen tanaman padi di Kabupaten Karawang | CART dan KNN | Lahan panen tanaman padi | Hasil dari perbandingan ini menunjukan bahwa algoritma CART bekerja lebih baik dari algortima KNN dengan dengan correlation coefficient 0,9646, MAE 498,6229, dan RMSE 834,0204 |
| 3 | Fakihatin Wafiyah, Nurul Hidayat, Rizal Setya Perdana | Implementasi Algoritma *Modified K-Nearest Neighbor* (MKNN) untuk Klasifikasi Penyakit Demam. | Kesulitan dalam mendiagnosa penyakit demam | *Modified K-Nearest Neighbor* | Penyakit demam | Hasil dari penerapan algoritma MKNN rata rata tingkat akurasi 92.42% |
| 4 | Ade Irpan Ramdani | Implementasi metode klasifikasi KNN pada sistem rekomendasi bidang keahlian mahasiswa: Studi kasus Teknik Informatika UIN Sunan Gunung Djati Bandung | memperbaiki penelitian sebelumnya yang dilakukan Muhammad Fikri Lazuardi yang memperoleh hasil yang hanya 68%. | K-nearest neughbor | Bidang keahlian mahasiswa | Hasil pengujian dalam penerapan algoritma KNN pada penelitian ini didapatkan tingkat akurasi 98,30% |
| 5 | Muhammad Faisal Amin | Penerapan Algoritma Cart Untuk Memprediksi Status Kelulusan Mahasiswa | Tingkat kelulusan prodi teknik informatika (TI) STMIK Banjarbaru sangat rendah, yaitu hanya 18,83%. | CART | mahasiwa | Berdasarkan hasil pengujian, akurasi dari model algoritma ini sebesar 64,37%[3]. |
| 6 | Anshori, Luthfi  Regasari, Rekyan  Putri,Mardi | Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya | Kesulitan dalam menentukan keminatan ini untuk perkuliahan mahasiswa yang nantinya juga berpengaruh pada tugas akhir (skripsi) hingga kelulusan mahasiswa | Mengimplementasikan metode KNN | Mahasiswa | Hasil dari pengujian 30 data maka dieroleh tingkat akurasi 76,66% terhadap data latih dengan nilai K yang paling optimal adalah K bernilai 10[7]. |
| 7 | Zumrotul Mutiah, Martaleli Bettiza,Muhamad Radzi Rathomi | Penerapan algoritma classification and regression tree (cart) untuk pemilihan jurusan sma  (studi kasus : sma negeri 3 bintan kabupaten bintan) | Kesulitan dalam pemilihan jurusan bagi siswa sekolah menengah atas (SMA) | Metode CART | Siswa sekolah menengah pertama (SMA) | Hasil dari keseluruhan pengujan hasil prediksi didaptkan sebesar 87%[8]. |
| 8 | Muhammad Reza Ravi, Indriati, Sigit Adinugroho | Implementasi Algoritma *Modified K-Nearest Neighbor* (MKNN) Untuk Mengidentifikasi Penyakit Gigi Dan Mulut | Mengidentifkasi jenis penyakit gigi dan mulut | *Modified K-nearest neighbor* | Penyakit gigi dan mulut | Hasil dari penelitian dengan data latih 70 dan daa uji 30 serta nilai k=6 didapatkan hasil 86,6%[9]. |

## **Beasiswa Ilmu Dasar Agama**

Beasiswa Ilmu Dasar Agama atau dikenal secara umum dengan Beasiswa Kajian Keislaman adalah beasiswa dari pemerintah departemen agama yang disediakan oleh Universitas Islam Negeri Sunan Gunung Djati Bandung, untuk alumni MA/SMA dan yang sederajat dimana bertujuan memperkuat kajian keislaman yang menjadi misi utama perguruan tinggi islam. Biaya yang diperoleh dari beasiswa IDA ini sebesar Rp.5.000.0000. Dengan adanya Beasiswa IDA diharapkan dapat mengurangi beban mahasiswa dalam belajar terutama dalam hal biaya.

Tujuan dengan adanya program ini adalah memberikan beasiswa kepada calon mahasiswa yang memiliki minat tinggi dalam memperdalam kajian keislaman serta menerapkan nilai ilmu keislaman melalui program Beasiswa IDA. Sasaran dari program beasiswa ini diharapkan dapat menghasilkan lulusan atau sarjana yang memiliki kualifikasi sebagai berikut:

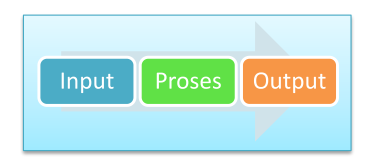
1. Penguasaan dasar dasar ilmu keislaman yang secara baik
2. Terampil berkomunikasi minimal dua bahasa (Arab dan Inggris) lisan dan tulisan serta hafal Al Quran.
3. Prodi tertentu dari 4 (empat) Fakultas (Ushuluddin, Dakwah, dan Komunikasi, Syari’ah dan Hukum, Adab dan Humaniora) yang mengalami kekurangan peminat.
4. Program dilaksanan dengan memberikan bimbingan khusus kepada mahasiswa dalam memperdalam ilmu keislaman.

Proses seleksi dilakukan berdasarkan data pelamar yang melakukan daftar bidikmisi namun tidak terjaring, dan yang memiliki prestasi rata rata nilai ijazah minimal sebesar 77 dan nilai rata rata UN sebesar 45. Dalam mekanisme seleksi peminat mendaftar secara manual dengan mendaftarkan diri kepada Wakil Dekan III Bidang Kemahasiswaan Fakultas penyelenggaraan program ,yang selanjutnya mengumpulkan berkas yang diperlukan. Sedangkan dalam melakukan proses seleksi penerimaan beasiswa IDA dilakukan secara manual, melalui Wakil Dekan III Fakultas penyelenggara program yang ditunjuk. Setelah dilakukan seleksi Wakil Dekan III menentukan orang yang berhak untuk mendapatkan beasiswa IDA sesuai dengan kuota masing-masing fakultas dan setelah ada penentuan yang mendapatkan beasiswa tersebut, maka Wadek III memberikan ke bagian kemahasiswaan untuk di proses SK penetapan sebagai Penerima Beasiswa Ilmu Dasar Agama (IDA).

## **Data Mining**

Secara umum data mining dibagi menjadi yaitu data dan mining. Dimana data adalah sekumpulan informasi atau data yang terekam.Sedangkan mining merupakan sebuah proses penampungan. Maka dari itu data mining adalah penampungan informasi atau data yang menghasilkan *output* pengetahuan[10].

Proses input,proses,output menjadi bagian dari fase data mining yang membedakannya hanya yang menjadi inputan dalam data mining berupa himpunan data , sedangkan pada prosesnya berupa algoritma data mining dan *output* berupa pengetahuan[10].



Gambar 2.1 Proses input ke output

## **Klasifikasi**

Klasifikasi merupakan sebuah metode dalam penglompokan data sesuai dengan aturan yang ditetapkan, dimana mengelompokan data berdasarkan variael yang diteliti bertujuan memperkirakan objek yang belum diketahui kelasnya[7].

Tujuan dari klasifikasi yaitu untuk mengelompokan secara akurat data data yang belum diketahui kategori kelasnya[7].

## **Normalisasi Data**

Dalam data mining proses normalisasi data sangat diperlukan dimana bertujuan mempersempit *range* data training. Pada proses nomalisasi data ini terdapat beberapa metode seperti, min-max normalization , z-core normalization, decimal scalling dan sigmoidal normalization [1].

Pada penilitian ini akan menggunakan metode normalisasi data min-max normalization. Metode min-max normalization merupakan perubahan nilai dari data yang sudah ada dengan range value terkecil(min) sebesar 0 dan nilai terbesar(max) sebesar 1. Berikut persamaan min-max normalization[1].

(1)

Keterangan :

V’ = Nilai dari data baru hasil dari normalisasi

V = Nilai dari data sebelum dinormalisasi

= Batas nilai maksimum yang baru

= Batas nilai minimum yang terbaru

Max A = nilai maksimum pada kolom

Min A = nilai minimum pada kolom

## ***K-nearest Neighbor***

*Metode k-nearest neighbor*  merupakan metode klasifikasi yang sering digunakan dalam memecahkan masalah data mining. Tujuan dari *k-nearest neighbor* yaitu mengklasifikasikan data berdasarkan jarak data training terdekat dengan data yang baru. Dalam *k-nearest neighbor* jumlah data ditentukan oleh pengguna dengan dinyatakan oleh nilai *k*[7]*.*

Berikut ini langkah langkah dalam metode KNN[11].

1. Menentukan nilai parameter *k.*
2. Menghitung jarak antar data yang akan dicari berdasarkan dat latih
3. Mengurutkan data jarak yang sudah terbentuk
4. Menentukan jarak terdekat hingga urutan *k.*
5. Mencari nilai jumlah kelas tetangga yang terdekat dan ditetapkan kelas nya sebagai data testing yang akan di cari.

(2)

Keterangan :

= data training

= data testing

= data training

= jarak

= dimensi data

## ***Modified K-nearest Neighbor***

*Modified k-nearest neigbor* (MKNN) merupakan sebuah ekspansi dari metode *K-nearest neighbour* dimana dilakukan proses perhitungan data training dengan tetangga terdekatnya, dan hasil dari perhitungan data training diklasifikasikan dnegan data uji yang akan dicari. Perbedaan dari *modified k-nearest neighbor* dan *k-nearest neighbor* adalah proses perhitungannya. Dimana dalam metode mknn ini terdapat proses perhitungan validitas data dan proses *weight voting* sedangkan metode knn tidak melakukan perhitungan validitas data dan *weight voting*[9].

### Jarak *Eucliddean*

Langkah dalam menghitung jarak *euclidian* yaitu dengan menentukan terlebih dahulu data training . Setelah data training dipatkan maka tahap selanjutnya melakukan proses perhitungan untuk mencari jarak terdekatnya dengan *k*, dimana tujuan validitas data adalah melihat jumlah titik dengan table yang sama terhadap semua data training, dimana pada tahap ini nilai fungsi *s* bergantung pada nilai data training tetangga terdekatnya[1].

Persamaan dalam tahap validitas data sebagai berikut.

(3)

Keterangan :

K = Jumlah titik terdekat

Label(x) = label kelas x

Ni(x) = Label kelas titik terdekat x

S = bernilai 1 apabila kelas sama atau bernilai 0 apabila kelasnya tidak sama

### *Weight Voting*

Proses penting selanjutnya adalah proses *weight voting* , dimana pada proses ini berguna untuk data training dengan nilai validitas tinggi serta jarak terdekat dengan data uji. Dimana langkah pertama yaitu menghitung bobot setiap tetangga yang selanjutnya, validitas sampel data training dikali dengan nilai bobot *euclidean* tersebut, berikut ini persamaan *weight voting* [1].

(4)

Keterangan :

W(i) = Perhitungan *weight voting* ke i

Validitas(i) = nilai validitas ke i

= Jarak *euclidean* data training dan data uji

α= Nilai alpha

## ***Clasification and Regression Trees***

Metode klasifikasi classification and regression tress (CART) dapat dipakai pada datayang berjumlah besar, yaitu dimana data dengan banyak variabel atau data dengan variabel campuran berdasarkan pemilihan biner[2]. Pada umumnya teknik yang digunakan dalam metode CART ini menggunakan teknik pohon keputusan,sehingga metode ini dapat dikatakan sederhana tetapi memiliki hasil lebih mudah diintrepetasikan, akurat dan perhitungannya yang lebih cepat apabila dikomparasikan dengan metode klasifikasi lainnya[3].

Tujuan dari metode CART ini untuk memperoleh suatu kumpulan data yang akurat sebagai salah satu ciri pengklasifikasian. Metode CART ini memiliki salah satu ciri yaitu berupa *node* keputusan yang selalu bercabang dua atau biner. Pada metode ini sebuah *record* diklasifikasikan dala m salah satu klasifikasi yang tersedia terhadap variabel tujuan berlatarkan nilai variabel prediktornya. Berikut langkah langkah algoritma CART[8]:

1. Menyusun secara lengkap calon cabang (*candidate split*) yang dilakukan terhadap seluruh variabel prediktornya. Daftar yang berisi calon cabang disebut calon cabang mutakhir
2. Beri nilai keseluruhan pada calon cabang mutakhir dengan menghitung nilai besaran kesesuaian Ø (s/t).
3. Tentukan calon cabang yang memiliki kesesuaian Ø (s/t). Setelah noktah keputusan tidak ada lagi, maka algoritma CART dihentikan.

Ø (s\t ) = 2 Q (s/t) (1)

(2)

Keterangan :

= cabang kiri dari noktah keputusan t

= cabang kanan dari noktah keputusan t

(3)

(4)

(5)

(6)