**BAB II**

**TINJAUAN PUSTAKA DAN DASAR TEORI**

* 1. **Tinjauan Pustaka**

Penelitian yang mengimplementasikan *k-NN* pernah dilakukan oleh Risman (2012) yang membahas mengenai penentuan penerima beasiswa mahasiswa di STMIK Sinar Nusantara Surakarta. Data *input* berupa nomer induk mahasiswa, nama mahasiswa, jenis kelamin, alamat, dan kriteria beasiswa yang digunakan untuk penetuan (semester, ipk, penghasilan orang tua dan tanggungan orang tua). Bahasa pemrograman yang digunakan adalah *PHP*. Hasil yang dicapai yaitu dapat menetukan penerima beasiswa mahasiswa dengan tingkat keakuratan 90,90 %.

Rohman (2013) melakukan penelitian dengan metode *k-NN* untuk memprediksi kelulusan mahasiswa. Data yang dikumpulkan adalah data mahasiswa dengan program studi strata satu (S1) untuk tahun kelulusan 2011. Jumlah data yang digunakan sebanyak 1.633 data, dengan atribut nim, nama, usia, fakultas, dan IP semester satu sampai delapan. Model yang dihasilkan diuji untuk mendapatkan nilai *accuracy* dan *AUC (Area Under Curve)*  dari algoritma klasifikasi *data mining* sehingga menghasilkan data mahasiswa dengan klastering data k=1 dengan menggunakan *k-NN* didapat nilai *accuracy* adalah 82,25% dan nilai *AUC* adalah 0.500, dengan *cluster* data *k=2 accuracy* adalah 79,45% dan nilai *AUC* adalah 0.826, dengan *cluster* data *k=3 accuracy* adalah 83,95% dan nilai *AUC* adalah 0.853, dengan *cluster* data *k=4 accuracy* adalah 82,62% dan nilai *AUC* adalah 0.874, dengan *cluster* data *k=5 accuracy* adalah 85,15% dan nilai *AUC* adalah 0.888

Krisandi (2013) melakukan penelitian dengan menerapkan algoritma *k-NN* dalam klasifikasi data hasil produksi kelapa sawit pada PT. Minamas Kecamatan Parindu. Data yang digunakan adalah data hasil produksi kelapa sawit dari 50 kelompok tani pada periode Juli – Desember 2011. Dari penelitian ini menghasilkan informasi mengenai perkiraan hasil produksi kelapa sawit dimasa mendatang, berkisar pada hubungan kesamaan hasil produksi antar kelompok tani berdasarkan *cluster-*nya masing – masing. Dengan demikian dapat diselidiki akibat dari perbedaan yang mencolok dari hasil produksi kelompok tani yang ada pada *cluster* tersebut.

Mariana (2015) menerapkan algoritma *k-NN* untuk deteksi penyakit kanker *serviks*. Penelitian ini diimplementasikan dengan bahasa pemorgraman *PHP*, *Appserv* sebagai *web server*, dan *MySQL* sebagai *database*. Data *input* berupa gejala – gejala yang dirasakan oleh pengguna. Sehingga dapat melakukan diagnosa secara dini terhadap penyakit yang diderita. Dengan adanya aplikasi ini maka pengguna dapat mencari informasi mengenai penyakit kanker *serviks* dan konsultasi keluhan gejala penyakit kanker *serviks*.

Herwanto (2015) mengimplementasikan metode *k-NN* yang membahas mengenai pengklasifikasian data status gizi balita dengan studi kasus di UPT Puskesmas Pracimantoro 1, Kecamatan Pracimantoro, Kabupaten Wonogiri. Jumlah data yang digunakan sebanyak 281 data. Atribut yang digunakan berupa umur (bulan), berat badan (kg), dan jenis kelamin. Sistem ini dibuat dengan menggunakan bahasa pemrograman java berbasis desktop. Database yang digunakan adalah *MySQL*. Dengan adanya aplikasi ini pengguna dapat mengetahui status gizi balita apakah gizi baik, gizi kurang atau gizi buruk dengan nilai k=3 dan tingkat kesamaan presentase keberhasilan sebesar 93%.

Untuk penelitian yang dibuat yaitu implementasi metode *k-NN* dalam mengklasifikasi gizi balita berbasis *web mobile*. Data yang digunakan berasal dari Posyandu Balita Kecamatan Klaten Tengah sebanyak 268 data periode 01 Februari 2013 – 18 Februari 2018. Data *input* berupa usia (bulan), berat badan (kg), dan jenis kelamin. Perbedaan dengan penelitian sebelumnya adalah teknologi yang digunakan dalam menyampaikan informasi. Aplikasi *desktop* dapat diakses pada lokal saja, sedangkan aplikasi *web mobile* dapat diakses kapan saja dan dimana saja.

**Tabel 2.1 Penelitian Dengan Metode *k-NN***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No.** | **Penulis** | **Objek Penelitian** | **Atribut** | **Fitur** |
| 1. | Risman, dkk, 2012. | Mahasiswa STMIK Sinar Nusantara Surakarta. | * Data semester mahasiswa. * Indeks prestasi kumulatif mahasiswa. * Penghasilan orang tua. * Tanggungan orang tua. | Menentukan penerima beasiswa mahasiswa. |
| 2. | Rohman, 2012. | Mahasiswa universitas dengan program setudi starta satu (S1) untuk tahun kelulusan periode 2011. | * Nim * Nama * Umur * Fakultas * Indeks presteasi semester * Indeks prestasi semester 1 sampai 8 | Memprediksi kelulusan mahasiswa. |

**Tabel 2.1 (Lanjutan)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No.** | **Penulis** | **Objek Penelitian** | **Atribut** | **Fitur** |
| 3. | Krisandi, dkk, 2013. | Data hasil produksi kelapa sawit pada PT. Minamas Kecamatan Parindu. | * Jumlah produksi   (Tonase). | Memperkirakan hasil produksi kelapa sawit dimasa mendatag. |
| 4. | Mariana, dkk, 2015. | Pasien yang mengalami gejala penyakit kanker serviks. | * Gejala * Hasil diagnosa | Mendeteksi penyakit (Kanker Serviks). Dan konsultasi keluhan gejala. |
| 5. | Herwanto, 2015. | Gizi Balita UPT Puskesmas Pacimantoro 1 | * Usia (bulan) * Berat badan (kg) * Jenis kelamin | Mengklasifikasi gizi balita. |
| 6. | Hafizh, 2018. | Gizi Balita Posyandu Klaten Tengah, Desa Jomboran, Klaten | * Usia (bulan) * Berat badan (kg) * Jenis kelamin | Mengklasifikasi gizi balita. |

* 1. **Dasar Teori**
     1. **Data Mining**

Secara sederhana *data mining* adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar (Davies, 2004).

*Data mining* juga disebut sebagai serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data (Pramudiono, 2007).

*Data mining* sering juga disebut *knowledge discovery in database* (KDD). *KDD* merupakan kegiatan yang meliputi pengumpulan data, pemakaian data, *historis* untuk menemukan keteraturan, dan pola atau hubungan dalam *set data* berukuran besar (Santoso, 2007).

* + 1. **Konsep Klasifikasi**

Klasifikasi merupakan suatu teknik dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan. Teknik ini dapat memberikan klasifikasi pada data baru dengan memanipulasi data yang ada yang telah diklasifikasikan dan dengan menggunakan hasilnya untuk memberikan sejumlah aturan. Aturan tersebut digunakan pada data - data baru untuk diklasifikasikan. Teknik ini menggunakan *supervised induction*, yang memanfaatkan kumpulan pengujian dari record yang terklasifikasi untuk menentukan kelas – kelas tambahan (Kusnawi, 2007).

* + 1. **Definisi *K-Nearest Neighbor***

*k-NN* adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Data pembelajaran diproyeksikan ke ruang berdimensi banyak, di mana masing – masing dimensi merepresentasikan fitur dari data. Ruang ini dibagi menjadi bagian – bagian berdasarkan klasifikasi data pembelajaran. Nilai *k* yang terbaik untuk algoritma ini tergantung pada data. Secara umum, nilai *k* yang tinggi akan mengurangi efek *noise* pada klasifikasi, tetapi membuat batasan antara setiap klasifikasi menjadi lebih kabur. Nilai *k* yang bagus dapat dipilih dengan optimasi parameter, misalnya dengan menggunakan *cross-validation*. Kasus khusus di mana klasifikasi diprediksikan berdasarkan data pembelajaran yang paling dekat (dengan kata lain , *k* = 1) disebut algoritma *nearest neighbor*.

* + 1. **Algoritma *K-Nearest Neighbor***

Langkah – langkah perhitungan *k-NN* yaitu (Kusumadewi, 2009) :

1. Menentukan nilai parameter k (jumlah tetangga terdekat).
2. Hitung jarak diantara data yang akan diuji dengan semua data pelatihan atau data sampel. Untuk menghitung jaraknya menggunakan rumus *Euclidean Distance*:

*2*............................................................... (2.1)

Keterangan :

= jarak variable ke-*i*

*i* = variabel data (i *= 1,2,3, …, n*)

*n =* dimensi data

*p* = data uji (*testing*)

*q* = data sampel (*training*)

1. Urutkan jarak dari yang terkecil kemudian tentukan jarak terdekat sampai urutan ke – *k*.
2. Carilah jumlah kategori terbanyak dari tetangga yang terdekat, dan tetapkan kategori tersebut sebagai kategori dari data yang diuji.

Contoh kasus, seorang balita laki-laki dengan usia 15 bulan dan berat badan 7 kg, status gizinya adalah ?

Berikut adalah sampel data atau data latih :

**Tabel 2.2 Sampel Data**

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Usia (bulan)** | **Berat Badan (kg)** | **Status Gizi** |
| 1 | 12 | 5.6 | Gizi Baik |
| 2 | 14 | 5.3 | Gizi Kurang |
| 3 | 14 | 6 | Gizi Baik |
| 4 | 15 | 8.6 | Gizi Lebih |
| 5 | 15 | 7.2 | Gizi Baik |
| 6 | 16 | 5.4 | Gizi Buruk |
| 7 | 17 | 11 | Gizi Lebih |
| 8 | 17 | 6.5 | Gizi Kurang |
| 9 | 18 | 6.2 | Gizi Kurang |
| 10 | 18 | 5 | Gizi Buruk |

Lakukan langkah-langkah cara perhitungan *k-NN* berdasarkan *Euclidean Distance*:

1. Menentukan parameter k :

berarti nilai *k* = 3

1. Hitung jarak setiap sampel data dengan data yang akan diuji :

*2*

Sehingga diperoleh hasil perhitungan jarak antara sampel data dengan data yang diuji :

**Tabel 2.3 Hasil Perhitungan Jarak**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No** | **Usia (bulan)** | **Berat Badan (kg)** | **Status Gizi** | **Jarak (*d*)** |
| 1 | 12 | 5.6 | Gizi Baik | 3.31 |
| 2 | 14 | 5.3 | Gizi Kurang | 1.97 |
| 3 | 14 | 6 | Gizi Baik | 1.41 |
| 4 | 15 | 8.6 | Gizi Lebih | 1.60 |
| 5 | 15 | 7.2 | Gizi Baik | 0.20 |
| 6 | 16 | 5.4 | Gizi Buruk | 1.88 |
| 7 | 17 | 11 | Gizi Lebih | 4.47 |
| 8 | 17 | 6.5 | Gizi Kurang | 2.06 |
| 9 | 18 | 6.2 | Gizi Kurang | 3.10 |
| 10 | 18 | 5 | Gizi Buruk | 3.60 |

1. Urutkan data berdasarkan jarak terkecil dan tentukan  
   jarak terdekat sampai urutan nilai *k* = 3, maka diambil 3 jarak terkecil.

**Tabel 2.4 Hasil Pengurutan Jarak**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No** | **Usia (bulan)** | **Berat Badan (kg)** | **Status Gizi** | **Jarak (*d*)** |
| 1 | 15 | 7.2 | Gizi Baik | 0,20 |
| 2 | 14 | 6 | Gizi Baik | 1,41 |
| 3 | 15 | 8.6 | Gizi Lebih | 1,60 |
| 4 | 16 | 5.4 | Gizi Buruk | 1,88 |
| 5 | 14 | 5.3 | Gizi Kurang | 1,97 |
| 6 | 17 | 6.5 | Gizi Kurang | 2,06 |
| 7 | 18 | 6.2 | Gizi Kurang | 3,10 |
| 8 | 12 | 5.6 | Gizi Baik | 3,31 |
| 9 | 18 | 5 | Gizi Buruk | 3,60 |
| 10 | 17 | 11 | Gizi Lebih | 4,47 |

1. Cari jumlah kategori terbanyak dari tetangga terdekat pada tabel 2.4, dan tetapkan kategori tersebut sebagai kategori dari data yang diuji.  
   Dari hasil pengambilan data sejumlah nilai *k* pada tabel 2.4, yaitu *k* = 3, maka terdapat 2 status gizi baik dan 1 status gizi lebih. Oleh karena itu untuk balita dengan usia 15 bulan dan berat badan 7 kg, maka status gizinya adalah “Gizi Baik”.
   * 1. **Status Gizi**

Istilah gizi dapat diartikan sebagai proses dari organisme dalam mengguankan bahan makanan melalui proses pencernaan, penyerapan, transportasi, penyimpanan, metebolisme, dan pembuangan yang dipergunakan untuk pemeliharaan hidup, pertumbuhan fungsi organ dan produksi serta menghasilkan energi (Supariasa, 2001).

Status gizi merupakan faktor yang terdapat dalam level individu (level yang paling mikro). Faktor yang mempengaruhi secara langsung adalah asupan makan dan infeksi. Pengaruh tidak langsung dari status gizi yaitu ketahanan pangan dalam keluarga, pola pengasuhan anak, dan lingkungan kesehatan yang tepat, termasuk akses terhadap pelayanan kesehatan (Riyadi, 2001).

Status gizi juga dapat diartikan sebagai keadaan fisik seseorang atau sekelompok orang yang ditentukan dengan salah satu atau kombinasi dari ukuran – ukuran gizi tertentu (Soekirman, 2000). Dari beberapa pengertian di atas, dapat disimpulkan bahwa status gizi merupakan suatu ukuran keseimbangan antara kebutuhan dan masukan nutrisi yang diindikasikan oleh variabel terntu (Supariasa, 2001).

* + 1. **Klasifikasi Status Gizi Balita**

Dalam menetukan status gizi balita harus memiliki ukuran baku yang disebut *reference*. Pengukuran baku antropometri yang sekarang digunakan di Indonesia adalah *WHO-NHCS*. Supariasa (2002) menyatakan, klasifikasi status gizi dapat dibedakan menjadi empat yaitu :

1. **Gizi Lebih**

Kelebihan berat badan pada balita terjadi karena ketidakmampuan antara energi yang masuk dengan keluar, terlalu banyak makan, terlalu sedikit olahraga atau keduanya. Kelebihan berat badan anak tidak boleh diturunkan, karena penyusutan berat akan sekaligus menghilangkan zat gizi yang diperlukan untuk pertumbuhan (Arisman, 2007).

1. **Gizi Baik**

Status gizi baik atau status gizi optimal terjadi bila tubuh memperoleh cukup zat-zat gizi yang digunakan secara efisien sehingga memungkinkan pertumbuhan fisik, perkembangan otak, kemampuan kerja dan kesehatan secara umum pada tingkat setinggi mungkin (Almatsier, 2005).

1. **Gizi Kurang**

Status gizi kurang terjadi bila tubuh mengalami kekurangan satu atau lebih zat-zat esensial (Almatsier, 2005).

1. **Gizi Buruk**

Gizi buruk adalah suatu kondisi di mana seseorang dinyatakan kekurangan nutrisi, atau dengan ungkapan lain status nutrisinya berada di bawah standar rata-rata. Nutrisi yang dimaksud bisa berupa protein, karbohidrat dan kalori. Di Indonesia, kasus KEP (Kurng Energi Protein) adalah salah satu masalah gizi utama yang banyak dijumpai pada balita (Lusa, 2009).

* + 1. **Antropometri**

Salah satu bidang keilmuan ergonomi adalah *anthropometri*yang berasal dari kata *Antro* yang berarti manusia dan *Metron*yang berarti ukuran. Definisi dinyatakan sebagai suatu studi yang menyangkut pengukuran dimensi tubuh manusia dan aplikasi rancangan yang menyangkut geometri fisik, massa, dan kekuatan tubuh (Sritomo, 2003).

* + 1. **Kartu Menuju Sehat (KMS)**

Kartu menuju sehat yang sering disebut juga dengan KMS merupakan suatu kartu atau alat penting yang digunakan untuk memantau pertumbuhan dan perkembangan anak. KMS yang ada untuk saat ini adalah KMS balita, yaitu kartu yang memuat pertumbuhan serta indikator perkembangan yang bermanfaat untuk mencatat dan memantau tumbuh kembang balita setiap bulannya. Mulai dari lahir sampai berusia 5 tahun atau 60 bulan. KMS dapat diartikan sebagai rapor kesehatan dan gizi balita (Depkes, 2010).

Secara umum, KMS berisi gambaran kurva berat badan terhadap umur untuk anak berusia 0 – 5 tahun, dan catatan yang penting untuk diperhatikan oleh petugas dan orang tua, seperti riwayat kelahiran anak, pemberian ASI dan makanan tambahan pemberian imunisasi dan vitamin A, penatalaksanaan diare di rumah, serta patokan sederhana tentang perkembangan psikomotorik anak (Depkes, 2010).

* + 1. **Web Mobile**

*Web mobile* adalah aplikasi akses *internet* menggunakan peralatan yang bersifat *mobile* berbasiskan *browser* yang bertujuan untuk mengakses layanan data secara *wireless*. *Web mobile* dapat diakses pada perangkat *mobile* seperti telepon seluler, PDA (*Personal Digital Assistant*) dan perangkat *portable* yang tersambung ke sebuah jaringan telekomunikasi seluler (Simarmata, 2010).

* + 1. **XAMPP**

*XAMPP* merupakan sebuah *software* yang berfungsi untuk menjalankan *website* berbasis *PHP* dan menggunakan pengolah data *MySQL* dikomputer local. *XAMPP* berperan sebagai *web server* pada *computer local*. *XAMPP* juga dapat disebut *Cpanel virtual server*, yang dapat membantu melakukan *preview* sehingga dapat memodifikasi *website* tanpa harus *online* atau terakses dengan *internet* (Wicaksono, 2008).

* + 1. **PHP (*Hypertext Preprocessor*)**

*PHP* merupakan salah satu bahasa pemograman yang bersifat *open source* yang bekerja pada sisi *server*. *PHP* banyak digunakan saat ini untuk pemograman situs *web* dinamis (termasuk *blog*) meskipun penggunaan untuk hal lain juga memungkinkan (Kadir, 2009).

* + 1. **Bootstrap**

*Bootstrap* merupakan *framework* untuk membangun desain *web* secara responsif. Artinya, tampilan web yang dibuat oleh *bootstrap* akan menyesuaikan ukuran layer dan *browser* yang kita gunakan baik di *desktop*, *tablet* atau pun *mobile device*. Dengan *bootstrap* kita juga bisa membangun *web* dinamis atau pun statis (Alatas, 2013).

* + 1. **MariaDB**

*Server MariaDB* adalah salah satu *server database* terpopuler di dunia. Ini dibuat oleh pengembang asli *MySQL* dan dijamin tetap *open source*. Pengguna terkenal termasuk *Wikipedia*, *WordPress.com* dan *Google*.

*MariaDB* mengubah data menjadi informasi terstruktur dalam beragam aplikasi, mulai dari perbankan hingga situs web. Ini adalah pengganti *drop-in* yang disempurnakan untuk *MySQL*. *MariaDB* digunakan karena cepat, terukur dan kuat, dengan ekosistem mesin penyimpanan yang kaya, plugin dan banyak alat lainnya membuatnya sangat serbaguna untuk berbagai macam kasus penggunaan.

*MariaDB* dikembangkan sebagai perangkat lunak *open source* dan sebagai *database* relasional, ia menyediakan antarmuka *SQL* untuk mengakses data. Versi terbaru *MariaDB* juga menyertakan fitur *GIS (Geographic Information System)* dan *JSON (JavaScript Object Notation)* (https://mariadb.org/about/).