

BAB II

TINJAUAN PUSTAKA

2.1 Kemiskinan

Kemiskinan umumnya dilukiskan sebagai rendahnya pendapatan untuk memenuhi kebutuhan pokok. Di Indonesia pengukuran kemiskinan menggunakan kriteria dari BPS. BPS menentukan kriteria kemiskinan menggunakan pendekatan kebutuhan dasar (*basic needs*). Berdasarkan pendekatan kebutuhan dasar, ada 3 indikator kemiskinan yang digunakan, yaitu (1) *Headcount Index*, (2) indeks kedalaman kemiskinan (*Poverty Gap Index*), (3) indeks keparahan kemiskinan (*Poverty Severity Index*). *Headcount Index* digunakan untuk mengukur kebutuhan absolut yang terdiri dari dua komponen yaitu garis kemiskinan makanan (*food line*) dan garis kemiskinan non makanan (*non food line*). Garis kemiskinan BPS sebagai dasar untuk perhitungan *Headcount index* ditentukan berdasarkan batas pengeluaran minimum untuk konsumsi makanan setara dengan 2.100 kalori per hari dan konsumsi non makanan (Yacoup, 2012).

Menurut Sen menyatakan bahwa “kemiskinan jangan dianggap hanya sebagai pendapatan rendah (*low income*), tetapi harus dianggap sebagai ketidakmampuan kapabilitas (*capability handicap*)” (Yacoup, 2012).

Definisi yang umum digunakan adalah dari *World Bank*, dalam perhitungan dan kajian-kajian akademik pengertian kemiskinan adalah sebagai ketidakmampuan mencapai standar hidup minimum, sedangkan *United Nation Development Programs* mendefinisikan kemiskinan sebagai ketidakmampuan untuk memperluas pilihan-pilihan dalam hidup (Cahyat dkk, 2007). Kemudian BPS mendefinisikan kemiskinan sebagai ketidakmampuan untuk memenuhi standar minimum kebutuhan dasar yang meliputi kebutuhan makanan maupun non makanan. Inti dari model ini adalah membandingkan tingkat konsumsi penduduk dengan "garis kemiskinan" (GK) yaitu jumlah rupiah untuk konsumsi per orang per bulan (TNP2K, 2012).

Terdapat 14 kriteria keluarga sangat miskin berdasarkan Pendataan Program Perlindungan Sosial pada Tabel 2.1 yaitu:

Tabel 2.1 Kriteria Keluarga Sangat Miskin

| No | Kriteria | Keterangan |
|----|---|---------------|
| 1 | Luas lantai tempat tinggal kurang dari 8 meter per segi per orang. | Hampir Miskin |
| 2 | Lantai tempat tinggal terbuat dari tanah/bambu/kayu murahan. | |
| 3 | Jenis dinding tempat tinggal terbuat dari bambu/rumbia/kayu berkualitas rendah/tembok tanpa di plester. | |
| 4 | Tidak memiliki fasilitas buang air besar/bersama-sama dengan rumah tangga lain. | |
| 5 | Penerangan rumah tangga tidak menggunakan listrik. | |
| 6 | Sumber air minum berasal dari air sumur/mata air tidak terlindungi/sungai/air hujan. | |
| 7 | Bahan bakar memasak untuk sehari-hari adalah kayu bakar/arang/minyak tanah. | |
| 8 | Hanya mengonsumsi daging/susu/ayam satu kali dalam seminggu. | |
| 9 | Hanya membeli satu pasang pakaian baru dalam setahun | Miskin |
| 10 | Hanya sanggup makan sebanyak satu/dua kali dalam sehari | |
| 11 | Tidak sanggup membayar pengobatan di puskesmas/poliklinik. | |
| 12 | Sumber penghasilan kepala rumah tangga adalah petani dengan luas lahan 0,5 ha, buruh tani, nelayan, buruh bangunan, buruh perkebunan, atau pekerjaan lainnya dengan pendapatan di bawah Rp 600.000 per bulan. | Sangat Miskin |

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.

Tabel 2.1 Kriteria Keluarga Sangat Miskin (lanjutan)

| No | Kriteria | Keterangan |
|----|---|---------------|
| 13 | Pendidikan tertinggi kepala rumah tangga tidak sekolah/tidak tamat SD/hanya SD. | Sangat Miskin |
| 14 | Tidak memiliki tabungan/barang yang mudah di jauh dengan nilai minimal Rp 500.000, seperti sepeda motor(kredit/non kredit), emas, ternak, kapal motor, atau barang modal lainnya. | |

(Sumber: BPS 2009 dalam UPPKH Dinas Sosial dan Pemakaman kotaPekanbaru)

2.2 Program Keluarga Harapan (PKH)

Program Keluarga Harapan mulai dilaksanakan di Indonesia pada tahun 2007 dan diharapkan dapat dilaksanakan secara berkesinambungan, setidaknya hingga tahun 2015. Tahun 2007 merupakan tahap awal pengembangan program atau tahap uji coba. Tujuan uji coba adalah untuk menguji berbagai instrumen yang diperlukan dalam pelaksanaan PKH seperti antara lain metode penentuan sasaran, verifikasi persyaratan, mekanisme pembayaran, dan pengaduan masyarakat (TNP2K, 2012).

PKH dijalankan sebagai pelaksanaan dari UU No. 40 tahun 2004 tentang jaminan sosial. UU No. 11 tahun 2009 tentang kesejahteraan sosial. Inpres No. 3 tahun 2010 tentang program pembangunan yang berkeadilan. Perpres No. 15 tahun 2010 tentang percepatan penanggulangan kemiskinan dan UU No. 39 tahun 1999 tentang hak asasi manusia. PKH terfokus pada dua komponen yang berkaitan dengan peningkatan kualitas sumber daya manusia yaitu pada bidang kesehatan dan pendidikan.

Pada tahun 2007 PKH ini akan dilakukan uji coba di 7 Provinsi dengan jumlah sasaran program sebanyak 500.000 RTSM. Ketujuh Provinsi tersebut adalah Sumatera Barat, DKI Jakarta, Jawa Barat, Jawa Timur, Sulawesi Utara, Gorontalo, dan Nusa Tenggara Timur. Apabila tahap uji coba ini berhasil, maka PKH akan dilaksanakan setidaknya sampai dengan tahun 2015. Hal ini sejalan dengan komitmen pencapaian *Millenium Development Goals*, mengingat sebagian

indikatornya juga diupayakan melalui PKH. Selama periode tersebut, target peserta secara bertahap akan ditingkatkan hingga mencakup seluruh RTSM dengan anak usia pendidikan dasar dan ibu hamil/nifas (Tlonaen dkk, 2014).

Kedudukan PKH merupakan bagian dari program-program penanggulangan kemiskinan lainnya. PKH berada di bawah koordinasi Tim Koordinasi Penanggulangan Kemiskinan (TKPK) baik di pusat maupun di daerah. Oleh sebab itu akan segera dibentuk Tim Pengendali PKH dalam TKPK agar terjadi koordinasi dan sinergi yang baik. PKH merupakan program lintas Kementerian dan Lembaga, karena aktor utamanya adalah dari Badan Perencanaan Pembangunan Nasional, Departemen Sosial, Departemen Kesehatan, Departemen Pendidikan Nasional, Departemen Agama, Departemen Komunikasi dan Informatika, dan Badan Pusat Statistik. Guna menyukseskan program tersebut, maka dibantu oleh Tim Tenaga ahli PKH dan konsultan *World Bank* (TNP2K, 2012).

PKH adalah suatu program yang memberikan bantuan tunai kepada Rumah Tangga Sangat Miskin (RTSM), jika mereka memenuhi persyaratan yang terkait dengan upaya peningkatan kualitas sumberdaya manusia, yaitu pendidikan dan kesehatan. Tujuan utama dari PKH adalah untuk mengurangi kemiskinan dan meningkatkan kualitas sumber daya manusia terutama pada kelompok masyarakat miskin. Tujuan tersebut sekaligus sebagai upaya mempercepat pencapaian target MDGs (TNP2K, 2012).

- a. Meningkatkan kondisi sosial ekonomi RTSM.
- b. Meningkatkan taraf pendidikan anak-anak RTSM.
- c. Meningkatkan status kesehatan dan gizi ibu hamil, ibu nifas, dan anak di bawah 6 tahun dari RTSM.
- d. Meningkatkan akses dan kualitas pelayanan pendidikan dan kesehatan, khususnya bagi RTSM.

Tujuan utama PKH adalah membantu mengurangi kemiskinan dengan cara meningkatkan kualitas sumberdaya manusia pada kelompok masyarakat sangat miskin. Dalam jangka pendek, bantuan ini membantu mengurangi beban pengeluaran RTSM, sedangkan untuk jangka panjang dengan mensyaratkan

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.

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

keluarga penerima untuk menyekolahkan anaknya, melakukan imunisasi balita, memeriksakan kandungan bagi ibu hamil, dan perbaikan gizi, diharapkan akan memutus rantai kemiskinan antar generasi (TNP2K, 2012).

Sasaran atau penerima bantuan PKH adalah RTSM yang memiliki anggota keluarga yang terdiri dari anak usia 0-15 tahun dan/atau ibu hamil/nifas dan berada pada lokasi terpilih. Penerima bantuan adalah ibu atau wanita dewasa yang mengurus anak pada rumah tangga yang bersangkutan (jika tidak ada ibu maka: nenek, tante/ bibi, atau kakak perempuan dapat menjadi penerima bantuan). Jadi pada kartu kepesertaan PKH pun akan tercantum nama ibu/wanita yang mengurus anak, bukan kepala rumah tangga sehingga orang yang harus dan berhak mengambil pembayaran adalah orang yang namanya tercantum di Kartu PKH (TNP2K, 2012).

Calon Penerima terpilih harus menandatangani persetujuan bahwa selama mereka menerima bantuan, mereka akan: (1) Menyekolahkan anak 7-15 tahun serta anak usia 16-18 tahun namun belum selesai pendidikan dasar 9 tahun wajib belajar, (2) Membawa anak usia 0-6 tahun ke fasilitas kesehatan sesuai dengan prosedur kesehatan PKH bagi anak, dan (3) Untuk ibu hamil, harus memeriksakan kesehatan diri dan janinnya ke fasilitas kesehatan sesuai dengan prosedur kesehatan PKH bagi ibu hamil (TNP2K, 2012).

Komponen pendidikan dalam PKH dikembangkan untuk meningkatkan angka partisipasi pendidikan dasar wajib 9 tahun serta upaya mengurangi angka pekerja anak pada keluarga yang sangat miskin. Anak penerima PKH yang berusia 7-18 tahun dan belum menyelesaikan program pendidikan dasar 9 tahun harus mendaftarkan diri di Sekolah formal atau non formal serta hadir sekurang-kurangnya 85% waktu tatap muka. Setiap anak peserta PKH berhak menerima bantuan selain PKH, baik itu program nasional maupun lokal. Bantuan PKH bukanlah pengganti program-program lainnya karenanya tidak cukup membantu pengeluaran lainnya seperti seragam, buku dan sebagainya. PKH merupakan bantuan agar orang tua dapat mengirim anak-anak ke Sekolah (TNP2K, 2012).

Besaran bantuan tunai untuk peserta PKH bervariasi tergantung jumlah anggota keluarga yang diperhitungkan dalam penerimaan bantuan, baik

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.

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

komponen kesehatan maupun pendidikan. Besaran bantuan ini di kemudian hari bisa berubah sesuai dengan kondisi keluarga saat itu atau bila peserta tidak dapat memenuhi syarat yang ditentukan. Rincian jumlah bantuan termuat pada Tabel 2.2 di bawah ini.

Tabel 2.2 Jumlah Bantuan

| Nominal Bantuan | Bantuan Tetap | Komponen | | | Keterangan |
|-----------------|---------------|---------------------|-----------|-----------|--|
| | | Bumil/Nifas /Balita | Anak SD | Anak SMP | |
| 800.000 | 300.000 | 0 | 500.000 | 0 | bila 1 anak SD |
| 1.300.000 | 300.000 | 1.000.000 | 0 | 0 | bila ada bumil/nifas/balita |
| | 300.000 | 0 | 1.000.000 | 0 | bila ada 2 anak SD |
| | 300.000 | 0 | 0 | 1.000.000 | bila ada 1 anak SMP |
| 1.800.000 | 300.000 | 1.000.000 | 500.000 | 0 | bila ada bumil/nifas/balita dan 1 anak SD |
| | 300.000 | 0 | 500.000 | 1.000.000 | bila 1 anak SD dan 1 Anak SMP |
| | 300.000 | 0 | 1.500.000 | 0 | bila 3 anak SD |
| 2.300.000 | 300.000 | 1.000.000 | 0 | 1.000.000 | bila ada bumil/nifas/balita dan 1 anak SMP |
| | 300.000 | 0 | 1.000.000 | 1.000.000 | bila 2 anak SD dan 1 anak SMP |
| | 300.000 | 1.000.000 | 1.000.000 | 0 | bila ada bumil/nifas/balita dan 2 anak SD |
| | 300.000 | 0 | 0 | 2.000.000 | bila ada 2 anak SMP |
| 2.800.000 | 300.000 | 1.000.000 | 500.000 | 1.000.000 | bila ada bumil/nifas/balita dan 1 anak SD dan 1 anak SMP |
| | 300.000 | 1.000.000 | 1.500.000 | 0 | bila ada bumil/nifas/balita dan 3 anak SD |
| | 300.000 | 0 | 1.500.000 | 1.000.000 | bila ada 3 anak SD dan 1 anak SMP |
| | 300.000 | 0 | 500.000 | 2.000.000 | bila 1 anak SD dan 2 anak SMP |

(Sumber: Buku Panduan TNP2K)

2.3 PKH Kota Pekanbaru

PKH Kota Pekanbaru merupakan bantuan langsung tunai kepada Rumah Tangga Sangat Miskin yang berada di Kota Pekanbaru. Unit yang bertanggung jawab dalam melaksanakan kegiatan PKH adalah Unit Pelaksana Program keluarga Harapan (UPPKH) yang berada di bawah naungan Dinas Sosial dan Pemukiman Kota Pekanbaru. Kantor Sekretariat UPPKH terletak di Jalan Datuk Setia Maharaja No. 6.

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.

Alamat kontak atau *email*: uppkhkotapekanbaru@gmail.com. PKH di Kota Pekanbaru telah dilaksanakan sejak tahun 2013 sampai sekarang. Saat ini Koordinator UPPKH Kota Pekanbaru dijabat oleh Bapak Jasman, S.Sos.

2.4 Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan *machine learning* untuk mengekstraksi mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terikat dari berbagai *database* besar (Larose, 2005).

Data mining adalah suatu proses menemukan hubungan yang berarti, pola dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan, dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika. Kemajuan luar biasa yang terus berlanjut dalam bidang *data mining* didorong oleh beberapa faktor antara lain (Larose, 2005):

1. Pertumbuhan yang cepat dalam kumpulan data.
2. Penyimpanan data dalam *data warehouse*, sehingga seluruh perusahaan memiliki akses ke dalam *database* yang handal.
3. Adanya peningkatan akses data melalui navigasi *web* dan internet.
4. Teknik kompetisi bisnis untuk meningkatkan penguasaan pasar dalam globalisasi ekonomi.
5. Perkembangan teknologi perangkat lunak untuk *data mining* (ketersediaan teknologi).
6. Perkembangan yang hebat dalam kemampuan komputasi dan pengembangan kapasitas media penyimpanan.

Dari definisi-definisi yang telah disampaikan, hal penting yang terkait dengan *data mining* adalah:

1. *Data mining* merupakan suatu proses otomatis terhadap data yang sudah ada.
2. Data yang akan diproses berupa data yang sangat besar.

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.

3. Tujuan dari *data mining* adalah mendapatkan hubungan atau pola yang mungkin memberikan indikasi yang bermanfaat.

2.4.1. Kelompok *Data Mining*

Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu (Larose, 2005):

1. Deskripsi

Terkadang peneliti dan analis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Sebagai contoh, petugas pengumpulan suara mungkin tidak dapat menemukan keterangan atau fakta bahwa siapa yang tidak cukup profesional akan sedikit didukung dalam pemilihan presiden. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

2. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik dari pada ke arah kategori. Model dibangun menggunakan *record* lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya pada peninjauan berikutnya, estimasi nilai dari variabel target dibuat berdasar nilai variabel prediksi. Sebagai contoh, akan dilakukan estimasi tekanan darah *systolic* dari pasien rumah sakit berdasarkan umur pasien, jenis kelamin, indeks berat badan dan level sodium darah. Hubungan antara tekanan darah *systolic* dan nilai variabel prediksi dalam proses pembelajaran akan menghasilkan model estimasi. Model estimasi yang dihasilkan dapat digunakan untuk kasus baru lainnya. Contoh lain yaitu estimasi nilai indeks prestasi kumulatif mahasiswa program pasca sarjana dengan melihat nilai indeks prestasi mahasiswa tersebut pada saat mengikuti program sarjana.

3. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang. Beberapa

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.

metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi.

4. Klasifikasi

Dalam klasifikasi, terdapat target variabel kategori. Contoh, penggolongan pendapatan, dapat dipisahkan dalam 3 kategori yaitu pendapatan tinggi, pendapatan sedang dan pendapatan rendah.

Contoh lain klasifikasi dalam bisnis dan penelitian adalah:

1. Menentukan apakah suatu transaksi kartu kredit merupakan transaksi yang curang atau bukan.
2. Memperkirakan apakah suatu pengajuan hipotek oleh nasabah merupakan suatu kredit yang baik atau buruk.
3. Mendiagnosa penyakit seorang pasien untuk mendapatkan termasuk kategori penyakit apa.

5. Pengklasteran

Pengklasteran merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas obyek-obyek yang memiliki kemiripan. Klaster adalah kumpulan dari *record* yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan *record-record* dalam *cluster* lain. Pengklasteran berbeda dengan klasifikasi yaitu tidak adanya variabel target dalam pengklasteran. Pengklasteran tidak mencoba untuk melakukan klasifikasi, mengestimasi atau memprediksi nilai dari variabel target. Akan tetapi algoritma pengklasteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan (homogen), yang mana kemiripan *record* dalam satu kelompok akan bernilai maksimal sedangkan kemiripan dengan *record* dalam kelompok lain akan bernilai minimal.

Contoh pengklasteran dalam bisnis dan penelitian adalah:

1. Mendapatkan kelompok-kelompok konsumen untuk target pemasaran dari suatu produk bagi perusahaan yang tidak memiliki dana pemasaran yang besar.

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.

2. Untuk tujuan audit akuntansi, yaitu melakukan pemisahan terhadap perilaku *financial* dalam baik dan mencurigakan.
3. Melakukan pengklasteran terhadap ekspresi dari gen, untuk mendapatkan kemiripan perilaku dari gen dalam jumlah besar.
6. Asosiasi

Tugas asosiasi dalam *data mining* adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut sebagai analisis keranjang belanja.

Contoh asosiasi dalam bisnis dan penelitian adalah:

1. Meneliti jumlah pelanggan dari perusahaan telekomunikasi seluler yang diharapkan untuk memberikan respon positif terhadap penawaran *upgrade* layanan yang diberikan.
2. Menemukan barang dalam *supermarket* yang dibeli secara bersamaan dan barang yang tidak pernah dibeli secara bersamaan.

2.4.2. Proses *Data Mining*

Istilah *data mining* dan *Knowledge Discovery in Database* (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Salah satu tahapan dalam keseluruhan proses KDD adalah *data mining*. Proses KDD secara garis besar dapat dijelaskan sebagai berikut (Fayaad, 1996):

a. *Data Cleaning*

Data-data yang tidak relevan itu juga lebih baik dibersihkan karena keberadaannya bisa mengurangi mutu atau akurasi dari hasil *data mining* merupakan istilah yang sering dipakai untuk menggambarkan tahapan ini. Pembersihan data juga akan mempengaruhi performa dari sistem *data mining* karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

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.

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

b. *Data Integration*

Data mining tidak hanya berasal dari suatu *database* tetapi juga berasal dari beberapa *database* atau *file* teks, integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan, dan sebagainya.

c. *Data Transformation*

Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *data mining*. Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

d. *Data Mining*

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam *data mining* sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

e. *Pattern Evolution*

Dalam tahap ini hasil dari teknik *data mining* berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa, ada beberapa alternatif yang dapat diambil seperti menjadikannya umpan balik untuk memperbaiki proses *data mining*, mencoba teknik *data mining* lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat. Ada beberapa teknik *data mining* yang menghasilkan hasil analisa berjumlah besar seperti analisa prediksi. Visualisasi hasil analisis akan sangat membantu untuk memudahkan pemahaman dari hasil *data mining*.

f. *Knowledge*

Tahap terakhir dari proses *data mining* adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisa yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak memahami *data*

mining. Karenanya presentasi hasil *data mining* dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam proses *data mining*.

2.5 *K-Nearest Neighbor* (KNN)

KNN termasuk kelompok *instance-based learning*. Algoritma ini juga merupakan salah satu teknik *lazy learning*. KNN dilakukan dengan mencari kelompok K objek dalam data *training* yang paling dekat (mirip) dengan objek pada data baru atau data *testing* (Agrawal, 2014).

Teknik KNN dengan melakukan langkah-langkah yaitu (Santoso, 2007), mulai *input*:

- a. Data *training*, label data *training* K, data *testing*.
- b. Untuk semua data *testing*, hitung jaraknya ke setiap data *training*.
- c. Tentukan K data *training* yang jaraknya paling dekat dengan data.
- d. *Testing*.
- e. Periksa label dari K data ini.
- f. Tentukan label yang frekuensinya paling banyak.
- g. Masukkan data *testing* ke kelas dengan frekuensi paling banyak.
- h. Berhenti.

Secara umum untuk mendefinisikan jarak antara dua objek x dan y , digunakan Rumus jarak *Euclidean* pada Persamaan 2.1 (Thakur, 2013).

$$d_{xy} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \dots\dots\dots (2.1)$$

Keterangan:

- d = Jarak kedekatan
- X = Data *testing*
- Y = Data *training*
- n = Jumlah atribut 1 sampai n

2.5.1. Tujuan Algoritma KNN

Tujuan algoritma ini adalah mengklasifikasikan obyek baru berdasarkan atribut dan *training sample* (Larose, 2005). Algoritma KNN adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data

pembelajaran yang jaraknya paling dekat dengan objek tersebut. Dalam hal ini jumlah data/tetangga terdekat ditentukan oleh *user* yang dinyatakan dengan *K* (Agrawal, 2014).

2.5.2. Kelebihan dan Kekurangan Algoritma KNN

Adapun kelebihan dari algoritma KNN adalah algoritma ini lebih efektif di data yang besar, tangguh, dan mampu menghasilkan data yang lebih akurat. Sedangkan kekurangan dari algoritma ini adalah pertama, perlu untuk menentukan nilai *K* yang optimal sehingga untuk menyatakan jumlah tetangga terdekatnya lebih mudah. Kedua, persoalan terpenting pada metode KNN adalah biaya komputasi karena dalam proses klasifikasi, seluruh data dipakai untuk menentukan kelas dari pola tes (Imandoust dan Bolandraffar, 2013).

2.6 Modified K-Nearest Neighbor (MKNN)

Algoritma MKNN adalah sebuah metode pengembangan dari metode KNN, namun terdapat beberapa proses tambahan atau pengembangan dari metode KNN (Parvin, 2008), adapun tahapan MKNN adalah sebagai berikut:

1. Melakukan Penghitungan Jarak

Metode KNN merupakan metode dengan klasifikasi berdasarkan objek terdekat dari objek yang yang dicari (Parvin dkk, 2008).

Euclidean Distance

Penghitungan jarak antara dua titik yaitu titik pada data *training* (*x*) dan titik pada data *testing* (*y*) maka digunakan Rumus *Euclidean*, seperti yang ditunjukkan pada Persamaan 2.1 (Tahkur, 2013):

$$d_{xy} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \dots\dots\dots (2.1)$$

Dengan Penjelasan:

- d = Jarak antara titik pada *data training* *x* dan titik *data testing* *y* yang akan diklasifikasi.
- i = Memrepresentasikan nilai atribut.
- n = Merupakan dimensi atribut.

2. Validitas Data Training

Validitas digunakan untuk menghitung jumlah titik pada seluruh data latih, tetangga terdekat pada setiap data akan mempengaruhi validitas data.

Persamaan 2.2 mencari validitas (Parvin dkk, 2008):

$$\text{Validitas} = \frac{1}{K} \sum_{i=1}^K S(|b|(x), |b|(N_{i(x)})) \dots \dots \dots (2.2)$$

Dengan Penjelasan:

K = Jumlah titik terdekat.

|b|(x) = Kelas x

$N_i(x)$ = Label kelas titik terdekat x

Fungsi S digunakan untuk menghitung kesamaan antara titik x dan data ke-i dari tetangga terdekat. Persamaan 2.3 sebagai berikut (Parvin dkk, 2008):

$$S(a,b) = \begin{cases} 1_{a=b} \\ 0_{a \neq b} \end{cases} \dots \dots \dots (2.3)$$

Dengan Penjelasan:

a = Kelas a pada data latih.

b = Kelas selain a pada data latih.

3. Weight Voting

Menentukan *weight voting* adalah dengan menggunakan validitas dari tiap data pada data *training* dikalikan dengan *weight* berdasarkan pada jarak *Euclidean*. Dalam metode MKNN, perhitungan *weight voting* tiap tetangga seperti pada Persamaan 2.4 di bawah ini (Parvin dkk, 2008):

$$W(x) = \text{Validitas}(x) \times \frac{1}{d_e + 0,5} \dots \dots \dots (2.4)$$

Dengan Penjelasan:

W(x) = Perhitungan *Weight Voting*

Validitas = Nilai Validitas

d_e = Jarak *Euclidean*

2.6.1. Tujuan Algoritma MKNN

Algoritma MKNN adalah merupakan algoritma penyempurnaan dari Algoritma KNN yaitu sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. tahapan algoritma MKNN ditambah perhitungan validitas dan *weight voting* (Parvin dkk, 2008).

2.6.1. Kelebihan dan Kekurangan Algoritma MKNN

Adapun kelebihan dari algoritma MKNN adalah algoritma ini lebih efektif di data yang besar dan mampu menghasilkan data yang lebih akurat dari algoritma KNN. Sedangkan kekurangan dari algoritma ini adalah pertama, nilai K yang masih bias (Parvin, 2008) hingga perlu untuk menentukan nilai K yang optimal sehingga untuk menyatakan jumlah tetangga terdekatnya lebih mudah. Kedua, persoalan terpenting pada algoritma MKNN adalah persoalan komputasi lebih besar (Mufroffin, 2014).

2.7 Cross Validation

Cross validation digunakan dalam rangka menemukan parameter terbaik dari satu model. Ini dilakukan dengan cara menguji besarnya *error* pada data *testing*. Dalam *cross validation*, data dibagi ke dalam K sampel *cross* dengan ukuran yang sama. Dari K *subset* data yang digunakan akan dipakai K-1 sampel sebagai data *training* dan 1 sampel sisanya untuk data *testing*. Selanjutnya dilakukan proses *training* dan *testing* kemudian dihitung rata-rata *error* (*error mean*). Setiap *running* akan ditemukan *error* untuk *data testing*, model yang memberikan rata-rata *error* terkecil dipilih menjadi metode terbaik. Pada *cross validation* memiliki ketidakpastian dalam menentukan uji *cross* (Last, 2006).

2.8 Confusion Matrix

Confusion matrix adalah *tool* yang digunakan untuk evaluasi model klasifikasi untuk memperkirakan objek yang benar atau salah. Sebuah *matrix* dari prediksi yang akan dibandingkan dengan kelas yang asli dari *input-an* atau dengan kata lain berisi informasi nilai aktual dan prediksi pada klasifikasi (Patil dan Sherekar, 2013). Untuk variabel perhitungan *confusion matrix* dapat dilihat pada Tabel 2.3.

Tabel 2.3 *Confusion Matrix* 2 Kelas

| Classification | Prediction Class | |
|----------------|-----------------------|-----------------------|
| | Class = Yes | Class = No |
| Class = Yes | a (true positive TP) | b (false negative FN) |
| Class = No | c (false positive FP) | d (true negative TN) |

(Sumber: *Confusion Matrix*)

Adapun perhitungan tingkat akurasi pada *Confusion Matrix* 2 kelas berdasarkan Persamaan 2.5 (Patil dan Sherekar, 2013) adalah:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} = \frac{A+D}{A+B+C+D} \dots\dots\dots (2.5)$$

2.9 Penelitian Terdahulu

Penelitian terdahulu adalah penelitian yang berhubungan dengan penelitian penulis saat ini, beberapa penelitian terdahulu dapat dilihat pada Tabel 2.4.

Tabel 2.4 Penelitian Terdahulu

| No | Nama Penulis | Judul | Hasil |
|----|---|---|--|
| 1. | Trio Haloman, Wayan Firdaus, Sutrisno | Implementasi <i>Modified K-Nearest Neighbor</i> dengan otomasi nilai K pada pengklasifikasi penyakit tanaman kedelai. | Mengklasifikasi penyakit pada tanaman kedelai yang menghasilkan nilai akurasi yaitu 98,83%. |
| 2 | Hamid Parvin, Hosein Alizadeth, Minaei Bidgoli. | <i>Modified K-Nearest Neighbor</i> | Perbandingan nilai akurasi yang diuji pada data <i>wine</i> yang diambil dari data <i>irish</i> dengan hasil perbandingan yaitu KNN adalah 79,86% dan MKNN adalah 86,65% |

Tabel 2.4 Penelitian Terdahulu (lanjutan)

| No | Nama Penulis | Judul | Hasil |
|----|--|---|---|
| 3 | Hilmy Ainul Yaqien | Sistem Pendukung Keputusan penerima program keluarga harapan menggunakan metode <i>Naive Bayes</i> . | Menghasilkan sistem pendukung keputusan dengan mengambil atribut penilaian yaitu jumlah anak SD, jumlah anak SMP, Jumlah Balita, Jumlah Ibu Hamil, Jumlah Ibu Nifas. |
| 4 | Aruna Sign, Smita Shukla Patel | <i>Applying Modified K-Nearest Neighbor To Detect Insider Threat in Collaborative Information Systems</i> | Merupakan salah satu permodifikasian algoritma yang sangat handal dan menjanjikan dibanding metode MKNN dalam proses pemecahan masalah deviasi. |
| 5 | Siti Muftrofin, Abidatul Izzah, Arrie Kurniawandhani, M.Nasrul | Optimasi Teknik Klasifikasi <i>Modified K-Nearest Neighbor</i> menggunakan Algoritma Genetika. | Kemampuan MKNN lebih handal dibanding dengan KNN dari segi nilai akurasi yaitu 100% dan 88,23%. Namun nilai K masih bias, maka digunakan algoritma genetika untuk mencari nilai K yang optimal. |

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.

Tabel 2.4 Penelitian Terdahulu (lanjutan)

| No | Nama Penulis | Judul | Hasil |
|----|---|---|--|
| 6 | Zahra Reza Mohammadi, Hamid Alinejad, Hosein Alizadeh | <i>An extended Modified K-Nearest Neighbor.</i> | Pengembangan algoritma KNN menghasilkan algoritma MKKN penambahan langkah <i>validity</i> dan <i>weight voting</i> . Nilai akurasi MKNN lebih tinggi dari KNN dengan data yang diuji adalah data <i>Monk 3</i> , <i>isodata</i> , <i>wine</i> , <i>iris</i> , <i>balance-sc</i> , <i>Bupa</i> , <i>SAHeart</i> . |

(Sumber: Penelitian Terdahulu)

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.