**Nama : Meilina**

**NIM :09031181621012**

**TUGAS MTI**

**KDD (KNOWLEDGE DISCOVERY IN DATABASE)**

**Studi Kasus pada Koperasi**

Dari 5 jurnal yang saya cari dan temukan, beberapa analisis dan kesimpulan yang saya dapatkan antara lain:

1. Jurnal pertama dengan judul “**METODE NON HIERARCHY ALGORITMA *K-MEANS* DALAM MENGELOMPOKKAN TINGKAT KELARISAN BARANG** (STUDI KASUS : KOPERASI KELUARGA BESAR SEMEN PADANG)”. Koperasi Keluarga Besar Semen Padang, yang bergerak dalam bidang penjualan barang-barang kebutuhan sehari-hari, memiliki Toserba yang menyediakan barang yang lengkap. Sehingga Pimpinan kesulitan untuk mengetahui barang mana yang lebih di minati dan banyak di beli. Maka perlu di identifikasi dan di kelompokkan produk apa saja yang diminati sehingga bisa menyusun faktor-faktor apa saja yang dapat menarik konsumen baru untuk membeli. Berdasarkan masalah tersebut digunakannlah salah satu teknik yang di kenal dalam Data Mining yaitu clustering. Pengertian clustering adalah pengelompokkan sejumlah data atau objek kedalam cluster (group) sehingga setiap cluster akan berisi data yang semirip mungkin dan berbeda dengan objek dalam cluster yang lainnya. Ada dua metode clustering yang kita kenal, yaitu Hierarchy dan Non Hierarchy. Metode Hierarchy terdiri dari complete linkage clustering, single linkage clustering, average linkage clustering dan centroid linkage clustering. Sedangkan metode Non Hierarchy terdiri dari k-means dan Fuzzy k-means. Cluster dalam kasus ini yaitu barang yang paling laris dan yang kurang laris. Dengan kedua cluster tersebut ditambah dengan atribut barang (kode barang, nama barang, total stok, dan stok akhir) maka dapat dilakukan Proses analisa terhadap tingkat kelarisan yang dikategorikan “Barang Laris” dan “Barang Kurang Laris” menggunakan Algoritma K-Means, dimana Algoritma tersebut digunakan untuk mengelompokkan barang berdasarkan tingkat kelarisannya yang diambil sebagai sampel berdasarkan parameternya.



1. Jurnal kedua dengan judul “**DATA MINING UNTUK MEMPREDIKSI JENIS TRANSAKSI NASABAH PADA KOPERASI SIMPAN PINJAM DENGAN ALGORITMA C4.5”.** Untuk meningkatkan kinerja operasional, koperasi berusaha untuk melayani kebutuhan nasbahnya. Koperasi harus memperliahatkan jenis transkasi nasabah baik yang menabung, meminjam ataupun yang akan menjadi anggota sehingga dapat meningkatkan efektivitas kinerja koperasi. Maka dari itu kurangnya memprediksi jenis transaksi nasabah yang ada pada koperasi ini membuat para pemimpin koperasi kesulitan memberikan pinjaman dan juga menerima anggota baru dalam koperasi. Pimpinan kesulitan mengetahui profesi nasabah yang paling banyak meminjam pada koperasi. Berdasarkan permasalahan yang ada, sangat penting untuk memprediksi jenis transaksi nasabah demi meningkatkan kinerja operasional koperasi untuk melayani kebutuhan nasabahnya sendiri. Metode yang digunakan adalah Algoritma C 4.5. Algoritma C4.5 adalah salah satu metode untuk membuat decision tree berdasarkan training data yang telah disediakan. Algoritma C4.5 merupakan pengembangan dari ID3. Beberapa pengembangan yang dilakukan pada C4.5 adalah sebagai antara lain bisa mengatasi missing value, bisa mengatasi continiu data, dan pruning. Analisis yang sedang berjalan akan menguraikan jenis transaksi nasabah selama ini serta jenis transaksi nasabah dikoperasi simpan pinjam. Pohon keputusan (decision tree) merupakan salah satu teknik terkenal dalam datang mining dan merupakan salah satu metode yang popular dalam menentukan keputusan suatu kasus. Metode ini tidak memerlukan proses pengolahan pengetahuan terlebih dahulu dan dapat menyelesaikan dengan sederhana kasus-kasus yang memiliki dimensi besar. Tugas yang paling umum diserahkan kepada pohon keputusan adalah kualifikasi. Dari set database kita bisa mengetahui apakah suatu nasabah merupakan nasabah yang baik atau dari riwatnya, seseorang berpeluang terkenal suatu penyakit tertentu berdasarkan riwayat dan lain-lain.

****

1. Jurnal ketiga dengan judul “**Pencarian Nasabah dengan Menggunakan Data Mining dan Algoritma C 4.5 Koperasi Maduma Subang”.** Dalam pemberian Kredit perlu menganalisa kebutuhan kreditur, maka yang harus diketahui terlebih dahulu adalah prinsip-prinsip yang perlu ditegakan dalam rangka pemberian Kredit. Hal-hal yang perlu diperhatikan dalam pemberian kredit bagi nasabah adalah *Prinsip 6 C’s* *Analysis* yaitu sebagai berikut: *Character* adalah keadaan watak dari nasabah, baik dalamkehidupan pribadi maupun dalam lingkungan usaha. Kegunaan dari penilaian terhadap karakter ini adalah untuk mengetahui sampai sejauh mana kemauan nasabah untuk memenuhi kewajibannya (*willingness to pay*) sesuai dengan perjanjian yang telah ditetapkan. Sebagai alat untuk memperoleh gambaran tentang karakter dari calon nasabah tersebut, dapat ditempuh melalui upaya antara lain:
2. Meneliti riwayat hidup calon nasabah;
3. Meneliti reputasi calon nasabah tersebut di lingkungan usahanya;
4. Meminta *bank to bank information* (Sistem Informasi Debitur);
5. Mencari informasi kepada asosiasi-asosiasi usaha dimana calon nasabah berada.
6. Mencari informasi apakah calon nasabah suka berjudi;
7. Mencari informasi apakah calon nasabah memiliki hobi berfoya-foya.

Berdasarkan latar belakang dan rumusan masalah diatas, maka penelitian ini bertujuan untuk mengantisipasi jumlah nasabah yang melakukan pembayaran melewati jatuh tempo yang sudah ditetapkan agar tidak terjadi kenaikan jumlah kredit macet yang berpotensi terjadinya kerugian pada pihak Koperasi. Tujuannya adalah agar petugas Analys Kredit dapat mengetahui dan memiliki standar untuk menentukan pemberian kredit kepada calon Nasabah sehingga dapat meminimalisir terjadinya kredit macet yang berdampak terjadinya kerugian bagi Koperasi atau pihak nasabah. Sama dengan jurnal yang saya dapat sebelumnya, pemecahan masalamenggunakan metode Algritma C 4.5. goritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan. Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami. Dan mereka juga dapat diekspresikan dalam bentuk bahasa basis data seperti *Structured Query Language* untuk mencari *record* padakategori tertentu.





1. Jurnal yang keempat dengan judul “**PENERAPAN DATA MINING UNTUK MEMPREDISI MINAT ANGGOTA TERHADAP PRODUK KOPERASI PTPN VII MUSILANDAS”.** Koperasi PTPN VII Musi Landas merupakan koperasi yang bergerak di bidang penjualan kebutuhan sehari-hari. Kendala yang dihadapi oleh Koperasi PTPN VII Musi Landas yaitu kesulitan dalam menentukan produk yang diminati oleh anggotanya. Apabila koperasi ini dapat memprediksi hal tersebut maka akan meminimalisir kerugian dan meningkatkan penjualan yang berdampak pada *income* kepada koperasi. Untuk itu akan dikembangkan sebuah penerapan *data mining* yang berfungsi untuk memprediksi minat produk yang diminati anggota. Teknik *data mining* yang diterapkan adalah Klasifikasi sedangkan metode klasifikasi yang digunakan adalah *Decision Tree* (pohon keputusan). Algoritma yang dipakai adalah algoritma C4.5 dan DTREG sebagai perangkat lunak untuk menghasilkan pohon keputusan.



1. Jurnal kelima dengan judul **“PENERAPAN DATA MINING DENGAN METODE NAIVE BAYES UNTUK MEMPREDIKSI KELAYAKAN PENGAJUAN KREDIT PADA KOPERASI RUKUN ARTHA SANTOSA JUWANA PATI”.** Adapun KSU Rukun Artha Santosa ini merupakan sebuah bidang usaha yang bergerak di bidang serba usaha, dibentuk untuk mengatasi permasalahan ekonomi yang dialami oleh kelompok usaha perikanan tangkap di Juwana. Salah satu unit kegiatan utama KSU Rukun Artha Santosa adalah memberikan kredit simpan pinjam. Pada tahapan evaluasi dalam proses pemberian kredit menjadi masalah yang sangat penting bagi koperasi RAS, karena pihak koperasi masih kesulitan untuk menentukan pemohon yang layak mendapatkan fasilitas kredit dan tidak beresiko menyebabkan kredit macet. Masalah seperti ini terjadi karena kurang akuratnya pihak pemberi kredit dalam penilaian terhadap kemampuan nasabah, sehingga mengakibatkan kesalahan dalam keputusan pemberian kredit yang berujung pada kemacetan kredit [6]. Oleh karena itu untuk mengatasi dan mengantisipasi hal tersebut, pihak koperasi dalam memberikan pinjaman perlu penerapan prinsip kehati-hatian supaya resiko yang timbul dari pemberian kredit kepada calon debitur tidak terlalu besar. Dalam menerapkan prinsip kehati-hatian ini koperasi perlu melakukan manajemen resiko dengan mengidentifikasi dan memprediksi kelayakan nasabah dengan baik sebelum memberikan pinjaman dengan cara memperhatikan data historis pinjaman nasabah [8]. Cara untuk menentukan prediksi kelayakan nasabah dalam membayar pinjaman kepada koperasi yaitu dengan menggunakan sebuah algoritma *data mining* dan untuk menghitungya diperlukan keahlian khusus yang mengerti *data mining* [2]. *Data mining* dianggap mampu dalam proses pencarian informasi, mendapatkan Sebuah solusidalam pengambilan keputusan dan menemukan pola tersembunyi dari suatu data kemudian diolah sehingga menjadi sebuah informasi yang berguna bagi seseorang atau organisasi [9]. Dalam data mining terdapat beberapa algoritma klasifikasi yang dapat digunakan untuk analisa kredit, salah satunya adalah dengan algoritma *Naïve* *Bayes*. Klasifikasi *Naïve Bayes* berjalansangat baik dibandingkan dengan model klasifikasi lainnya. Kejadian ini dibuktikan dengan penelitian yang dilakukan oleh Xhemali, J.Hinde dan G.Stone (2009) yang berjudul *“Naïve* *Bayes vs. Decision Tree vs. Neural Networks in the Clasification of Training Web Pages”*, dalam penelitiantersebut menjelaskan bahwa “*Naïve* *Bayes Classifier* mempunyai tingkatakurasi yang lebih baik dibanding model *classifier* lainnya”, dengan tingkat akurasi 95,20% dan nilai F-Measure lebih 97% [13]. Penelitian lain yang dilakukan oleh Evaristus dan Aryuni (2014) berjudul *“Comparative* *Study of Data Mining Model for Credit Card Application Scoring In Bank”*,melaporkan klasifikasi *Naïve Bayes* memiliki tingkat akurasi yang lebih baik dari pada klasifikasi *ID3,* yaitu dengan tingkat keakuratan NBC adalah 82% dan ID3 adalah 76% [14]. Penelitian lainnya yang dilakukan oleh Suryati (2013) berjudul *“Aplikasi Pendukung* *Keputusan Pengajuan Kredit Sepeda Motor”,* metode Naïve Bayes telahdapat digunakan untuk menentukan kelayakan kredit sepeda motor [15]. Maka dari itu penggunaan metode Naïve Bayes ini akan sangat tepat untuk membangun sebuah aplikasi data mining untuk memprediksi penentuan kelayakan pengajuan kredit pada pemohon yang akan melakukan pengajuan kredit, sehingga pihak koperasi atau bagian kredit akan tepat dalam memperkirakan diterima atau ditolaknya pemohon yang akan melakukan pengajuan kredit.



***Knowledge Discovery In Database(KDD)***merupakan metode untuk memperoleh pengetahuan dari database yang ada. Dalam database terdapat tabel - tabel yang saling berhubungan / berelasi. Hasil pengetahuan yang diperoleh dalam proses tersebut dapat digunakan sebagai basis pengetahuan *(knowledge base)* untuk keperluan pengambilan keputusan.

**

Istilah *Knowledge Discovery in Database* *(KDD)* dan data mining seringkali digunakansecara 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, dan salah satu tahapan dalam keseluruhan proses KDD adalah data mining. Proses KDD secara garis besar dapat dijelaskan sebagai berikut:

**

1. *Data Selection*

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam *Knowledge* *Discovery in Database(KDD)* dimulai. Datahasil seleksi yang akan digunakan untuk proses data mining, disimpan dalam suatu berkas terpisah dari basis data operasional.

1. *Pre-processing / Cleaning*

Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data yang menjadi fokus *Knowledge* *Discovery in Database (KDD).* Proses *cleaning* mencakup antara lain membuangduplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak. Juga dilakukan proses*enrichment*,yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk *Knowledge Discovery in* *Database (KDD)*, seperti data atau informasieksternal lainnya yang diperlukan.

1. *Transformation*

*Coding* adalah proses transformasi pada datayang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses *coding* dalam *Knowledge Discovery in Database (KDD)* merupakan proses kreatifdan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

**

1. *Data Mining*

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

1. *Interpretation / Evaluation*

Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses *Knowledge Diccovery in Database* *(KDD)* yang disebut *interpretation*. Tahap inimencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

Definisi Data Mining :

1. Mengekstrak atau “mining” pengetahuan dari kumpulan data yg sangat besar
2. Ekstraksi informasi yg berguna dari data, dimana sebelumnya tidak diharapkan, tidak dikenal & implisit
3. Eksplorasi & analisis, secara otomatis atau semi-otomatis dari sekumpulan data yg sangat besar untuk memperoleh pola2 data yg berarti
4. Proses analisis database yg besar secara semi-otomatis untk menemukan pola yang valid, baru, berguna dan dapat dipahami manusia



Data mining merupakan bagian dari proses Knowledge Discovery in Databases (KDD) – Proses transformasi data mentah menjadi informasi berguna. Dibawah ini adalah gambar proses kerja data mining. Pada dasarnya *data mining* terdiri dari :

1. *Predictive*, metode yang menggunakan beberapa variabel yang ada untuk memprediksi nilai masa depan (belum diketahui) dari variabel lain. Contoh : *classification, regression, biases/anomalies detection*.
2. *Descriptive*, metode yang mengungkapkan pola dalamdata, agar mudah diinterpretasikan oleh pengguna. Contoh : *clustering, association rules, sequential* *patterns.*

Jadi dapat disimpulkan melalui pengertian, tahapan, dan penerapan KDD pada studi kasus, dalam hal ini saya ambil pada koperasi. KDD adalah penerapan metode dari data mining. Proses ekstraksi data yang banyak, sehingga didapatkan pola yang bisa dijadikan sebuah acuan untuk mengambil keputusan dengan metode-metode saintifik. Seperti yang terdapat pada kasus pada jurnal yang saya temukan, menggunakan metode Algoritma C 4.5, metode *Clustering,* dimana terdapat rumus-rumus perhitungan berdasarkan data yang ada dan pola yang ditemuknn. Hasil perhitungan dianalisis sehingga bisa menjadi tolok ukur dala mengambil keputusan.