BAB II

TINJAUAN PUSTAKA

II.1 Klasifikasi

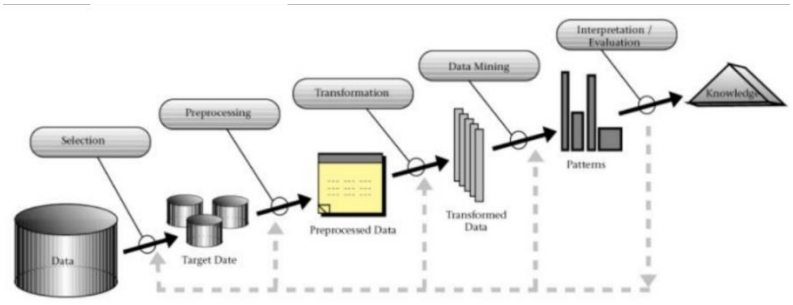
Klasifikasi dibentuk dari suatu teknik dari penambangan data yang bertujuan untuk membagi beberapa unsur *item set* data menjadi berbagai jenis kelas yang ada(Paquin et al., 2015). Klasifikasi merupakan pengorganisasian data dari objek data kelompok yang berbeda. Klasifikasi menetapkan setiap *instance* ke kelas tertentu sehingga kesalahan akan menjadi lebih sedikit, yang digunakan untuk mengekstraksi model secara akurat(Jadhav & Channe, 2016). Klasifikasi berperan besar dalam menghasilkan pohon keputusan yang tepat bedasarkan atribut yang digunakan. Penerapan klasifikasi harus mengidentifikasi atribut stok barang inventory penggunaan barang yang dikaitkan dengan profil tertentu tidak menjadi kerugian bagi perusahaan (Masankaran et al., 2018)

Proses klasifikasi menghasilkan model yang menyerupai gambar dan kelas atau konsep data yang menangani atribut numerik dan kategori. Klasifikasi memprediksi setiap label *class* kategori dan mengklasifikasikan data berdasarkan dari dataset pelatihan. Karakteristik dari klasifikasi adalah ketepatan *classifier* mengklassifikasi tuple secara akurat yang didasarkan karakteristik, waktu yang dibutuhkan untuk membangun model, kekuatan dari kemampuan mengklasifikasi *tuple* dengan memiliki *noise*, ukuran data harus independent dari ukuran basis data, fitur harus ditambahkan pada saat yang diperlukan (Gorade et al., 2017).

II.2 Data Mining

Data mining menurut(Rahim et al., 2018) adalah bidang ilmu multi disiplin, yang area kerja mengarah pada teknologi basis data, statistik, *machine learning*, pengenalan pola, jaringan saraf tiruan, pencarian informasi, sistem berbasis pengetahuan, komputasi kinerja tinggi, kecerdasan buatan, dan visualisasi data. Data mining diartikan sebagai penambangan data ataupun upaya untuk menemukan informasi yang berguna dan berharga pada database yang sangat besar. Hal terpenting dari teknik data mining adalah aturan menemukan pola frekuensi tinggi di antara set - set item yang disebut fungsi *Association Rules.* Data mining sering disebut sebagai *Knowledge Discovery in Database* (KDD) (Rahim et al., 2018).

KDD merupakan aktivitas yang mengekstraksi pola dari data. Pola itu ditemukan tergantung pada data mining yang diterapkan. Umumnya ada dua jenis tugas penambangan data yaitu tugas dari penambangan data deskriptif yang menghasilkan data yang bersifat umum, dan *data mining* yang bersifat prediktif tugas yang berusaha melakukan prediksi berdasarkan data yang telah ada. Data mining biasanya dilakukan pada data yang bersifat kuantitatif, multimedia, atau tekstual. Adapun alur proses dari KDD dapat dilihat dari gambar II.1 berikut ini (Sugara et al., 2018):



Gambar II.1. Alur Proses KDD

Sumber : Jayawardanu dan Hansun (2015) disitasi dalam (Sugara et al., 2018)

Adapun alur proses KDD memiliki 5 tahapan yang dapat dijelaskan sebagai berikut (Sugara et al., 2018) :

1. *Selection* memiliki tahapan untuk menyeleksi *multiple data source* yang memiliki data *noise* ataupun *missing value.*
2. *Prepocessing* memiliki tahapan untuk menggabungkan seluruh *data source* yang telah terkumpul sebelum dilanjutkan ke tahapan berikutnya.
3. *Transformasi* memiliki tahapan untuk mengubah data kedalam bentuk yang lebih sesuai dengan data mining.
4. Data Mining memiliki tahapan untuk membuat pola dari data yang telah dikumpulkan.
5. *Interpretation/Evaluation* memiliki tahapan yang dapat dilakukan untuk menginterprestasikan dan mengevaluasi pola yang telah ditemukan dan didefiniskan *knowledge* terhadap pola yang telah diterapkan.

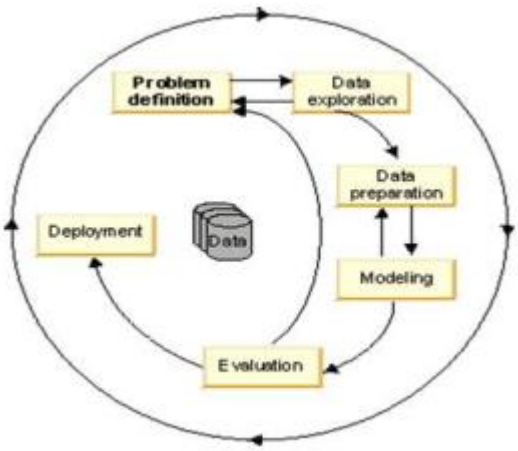
Data mining merupakan tahapan penggalian data yang memiliki beberapa langkah kunci yang dijelaskan sebagai berikut (Sugara et al., 2018) :

1. *Problem definition* memiliki tujuan untuk menentukan rangkaian permasalahan untuk membangun sebuah model yang digunakan.
2. Data *exploration* merupakan teknik pengumpulan data dan penyimpanan data yang semua data dikonsolidasikan sehingga bisa semua data konsisten.
3. Data *preparation* memiliki tujuan untuk mengubah data sehingga nilai yang dihasilkan valid dan konsisten sehingga analisis data lebih akurat.
4. *Modeling* merupakan isi dari kombinasi algoritma yang di analisis

didalamnya termasuk teknik statistik dan *clustering* tetapi juga termasuk penerus teknik seperti *decision tree*, jaringan dan algoritma yang berbasis aturan untuk menghasilkan tujuan tertentu.

5. *Evaluation and Deployment* dilakukan untuk menentukan kesimpulan utama analisis dan buat serangkaian rekomendasi untuk di pertimbangkan

Berikut penjelasan yang telah di paparkan diatas dapat dilihat dari gambar II.2 sebagai berikut:



Gambar II.2. Tahapan *Data Mining*

Sumber : Krishnaiah, Narmisha dan Chandra (2013)

disitasi dalam (Sugara et al., 2018)

II.3 Algortima Naïve Bayes

Naïve Bayes adalah algoritma yang memanfaatkan proposisi bayes untuk memperkirakan bagaimana tingkat kepastian pada suatu skema berubah dengan bukti dengan menggunakan interpolasi bayesian dengan tingkat kepercayaan. Antisipasi dari tentang adanya hubungan faktual mengenai data tersebut dapat

dinyatakan dalam distribusi probabilitas atas hipotesis yang mendefinisikan hubungan antar data (Janani et al., 2019).

Naïve Bayes menurut (Krichene, 2017) merupakan salah satu jenis dari klasifikasi bayesian yang sebenarnya dikenal sebagai probabilitas sederhana dan efektif. Klasifikasi Naïve Bayes yang sebenarnya dikenal sebagai probabilitas yang menerapkan teorema bayes dengan asumsi independen yang kuat (*naif)*. Klasifikasi Naïve Bayes cocok dengan dimensi inputan data yang tinggi. Bayes sering sekali dapat mengungguli kelas yang lebih baik dari algoritma klasifikasi lainnya.

Naïve Bayes menurut penulis adalah sebuah teknik klasifikasi dari perhitungan probabilitas yang sederhana untuk menghitung sekumpulan probabilitas dengan menjumlahkan kombinasi dan frekuensi nilai dari satu *dataset* yang digunakan. Algoritma yang diterapkan merupakan teorema bayes dengan mengasumsikan dari beberapa atribut yang independen atau tidak terikat yang diberikan oleh nilai pada atribut di tiap kelas.

Naïve Bayes lebih sering digunakan pada data yang relatif besar dan dapat menangani data yang dianggap tidak lengkap serta kuat terhadap atribut yang tidak relevan pada data. Pada algoritma Naïve Bayes memiliki kelemahan berdasarkan nilai akurasi yang dipengaruhi pada jumlah atribut. Keuntungan penggunaan Naïve Bayes adalah bahwa algoritma ini hanya butuh beberapa jumlah data latih (data *training*) yang kecil untuk menentukan estimasi paremeter yang diperlukan dalam proses pengklasifikasian.

Dalam algoritma ini perhitungan ditunjukkan melalui persamaan II.1 (Janani et al., 2019) :

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \quad \text{(II.1)}$$

Di mana :

X : *Class* data yang belum diketahui.

H : Hipotesis merupakan suatu kelas spesifik

P(H|X) : Probabilitas hipotesis nilai H berdasarkan kondisi nilai X (posteriori probabilitas)

P(H) : Probabilitas hipotesis H (prior probabilitas)

P(X|H) : Probabilitas nalai X dari kondisi pada hipotesis nilai H

P(X) : Probabilitas nilai X

II.4 Algoritma *Decision Tree*

Algoritma *decision tree* adalah pohon yang digunakan untuk mewakili objek dalam bentuk nilai *boolean* yang terdiri dari node internal dan daun. Node *internal* menunjukkan sentiment objek yang disebut dengan atribut. Setiap node memiliki dua cabang yaitu cabang kanan mewakili tidak adanya sentimen objek dan cabang kiri mewakili keberadaan sentimen objek(Wotawa et al., 2019). Daun mewakili kepuasan pengguna dengan salah satu nilai "Ya" atau "Tidak". Nilai daun "Ya" Jika stok barang ditambah dan nilai daun "Tidak" jika stok barang tidak ditambah.

Algoritma *decision tree* terdiri dari beberapa objek untuk mengumpulkan model pohon keputusan yang mengarah pada pada perhitungan probabilitas dari tiap *record* terhadap kategori tersebut. C4.5 merupakan pengembangan dari ID3. Ada beberapa model algoritma yang dipakai dalam *decision tree* antara lain adalah

ID3, C4.5 dan CART (Kozak & Juszczuk, 2018).

II.5 Algoritma C4.5

Algoritma C4.5 merupakan algoritma yang menghubungkan dengan *decision tree* yang mengubah kumpulan data menjadi pohon keputusan dan aturan keputusan. C4.5 adalah algoritma yang cocok untuk masalah klasifikasi dan data mining C4.5 memetakan nilai atribut ke dalam *class* yang dapat diterapkan pada klasifikasi baru (Berhane et al., 2018).

Algoritma C4.5 menggunakan konsep dari *information gain* atau *entropy reduction* untuk memilih pembagian yang optimal. Tahapan dalam membuat pohon keputusan dengan cara sebagai berikut (Alkhalifi et al., 2020):

1. Mempersiapkan data *training* yang diambil dari data yang telah ada sebelumnya dan sudah dikelompokkan dalam beberapa kelas tertentu.
2. Menentukan akar pohon yang dihitung dari nilai *gain* yang tertinggi dari masing – masing atribut atau berdasarkan nilai *index entropy* terendah. Yang dihitung dengan nilai rumus persamaan berikut ini :

$Entropy(S) = \sum_{i=0}^{n} -pi * log_2 pi$ …… (II.2)

Rumus (II.2) merupakan perhitungan dari *entropy* yang digunakan untuk menentukan seberapa informatif atribut tersebut(Alkhalifi et al., 2020). Berikut keterangannya:

S : Himpunan Jumlah Kasus

n : Jumlah dari besarnya partisi S

pi : Jumlah kasus terhadap partisi ke-i

$Gain(S, A) = Entropy\ (S) - \sum_{i=1}^{n} \frac{|Si|}{|S|} * Entropy(Si)$ ..... (II.3)

Rumus (II.3) merupakan rumus yang digunakan dalam perhitungan *gain* setelah melakukan perhitungan *gain* setelah melakukan perhitungan *entropy*(Alkhalifi et al., 2020). Berikut keterangannya :

S :Himpunan dari Kasus

n :Jumlah dari partisi A

|*Si*| :Jumlah kasus dari partisi ke-*i*

|*S*| :Jumlah dari kasus *S*

II.6 Inventory

Klasifikasi inventory digunakan untuk membagikan kedalam beberapa kelas atribut yang dipisahkan berdasarkan data pengadaan barang inventory setiap bulan yang di rekap pada setiap akhir bulan jumlah pemakaian barang tersebut pada setiap departemen oleh bagian pegawai inventory. Sistem pencatatan dilakukan secara langsung untuk mendata kebutuhan setiap departemen yang ada. Klasifikasi tersebut dibagi dengan atribut pada inventory berupa klasifikasi barang, jenis barang, jumlah pesanan, stok barang, waktu permintaan, waktu pengadaan, durasi pemakaian, departemen.

Proses data mining dilakukan untuk menggali informasi dari data inventory yang sering di pesan oleh setiap departemen. Teknik klasifikasi dari inventory tersebut dilakukan dengan menggunakan algoritma Naïve Bayes dan C4.5 untuk mencari nilai dari probabilitas dan nilai *gain* tertinggi. Klasifikasi yang telah didapatkan akan menghasilkan model yang akan diuji dengan data yang telah ada

dan melakukan perhitungan dan menghasilkan nilai dari dari akurasi kedua algoritma tersebut dan menemukan hasil algoritma yang paling baik dan efisien yang akan digunakan pada pengadaan barang inventory.

II.7 *Waikato Environment for Knowledge Analysis* (WEKA)

Waikato environment for knowledge analysis (WEKA) adalah sebuah aplikasi open source yang digunakan untuk menghasilkan perhitungan algoritma pada data mining(Jadhav & Channe, 2016). WEKA memiliki GUI yang bersifat mudah untuk dipelajari dan di mengerti. Algoritma tersebut diterapkan ke data set sehingga menghasil nilai dari algoritma yang dibutuhkan. WEKA berisi alat pre-processing, klasifikasi, regresi, clustering, aturan asosiasi, dan visualisasi. Jadi, dari penelitian ini untuk pengklasifikasiannya menggunkan WEKA versi 3.8 yang menghasilkan nilai dari algoritma Naïve Bayes dan C4.5. Adapun penelitian terkait untuk mengetahui aplikasi tools WEKA adalah sebagai berikut :

Tabel II.1 Penelitian Terkait dengan *Tools* WEKA

No	Nama Penulis	Publikasi	Judul	Pembahasan
1	Ade Surya Budiman, Xanty Adhi Parandani	Jurnal SIMETRIS, Vol. 9 No. 1 April 2018	Uji Akurasi Klasifikasi dan Validasi Data Pada Penggunaan Metode Membership Function dan Algoritma C4.5 Dalam Penilaian Penerimaan Beasiswa	Perhitungan dan pengujian *Mean Square Error* (MAE) dan *Root Mean Square Erro* (RMSE) lebih mudah dipelajari dan mudah untuk dipahami dan Pengujian Classified Instances dapat dilihat dari *correctly* data terhadap seleksi Mahasiswa.

No	Nama Penulis	Publikasi	Judul	Pembahasan
2	Sayali D. Jadhav, H.P. Channe	International Journal of Science and Research (IJSR) , Vol. 5 Issue 1, January 2016	Comparative Study of K-NN, Naïve Bayes dan Decision Tree Classification Techniques	Perhitungan kuantitatif dan data disktit terhadap titik noise dengan rata - rata variansi dari atribut untuk klasifikasi yang memiliki estimasi peluang dengan tidak mengikut sertakan probabilitas berangka nol. Jika nilai nol di masukkan kedalam data probabilitas akan mengansumsikan atribut bebas dengan prediksi nilai nol pada aplikasi WEKA.
3	Khairul Zaman	UPI YPTK Jurnal KemTekInfo Vol. 3, No. 2, Desember 2016, Hal. 12-24	Penerapan Data Mining Menggunakan Algoritma C4.5 Untuk Menentukan Kelayakan Penerimaan Bantuan Rehabilitas Sosial Rumah Tidak Layak Huni (Studi Kasus Di Pemerintahan Kabupaten Solok Selatan)	Menghitung entropy dan gain dengan lebih simple sehingga menghasilkan pohon perhitungan dengan melihat visualize trees J48 yang memperlihatkan pohon keputusan dengan lebih simple.
4	M. Rahmadi, Fazrlyanor, Tuti Susanti	JURIKOM (Junal Riset Komputer), Vol 7 No.1, Februrari 2020	Uji Akurasi Dataset Pasien Pasca Operasi Menggunakan Algoritma Naïve Bayes Menggunakan *Tools* WEKA	Explorer yang digunakan untuk menggali lebih jauh dengan aplikasi WEKA, experimenter digunakan untuk melakukan pengujian statistik skema belajar dan kowledge flow dengan pengetahuan pendukung
5	Mr. Sudhir M.Gorade, Prof. Ankit Deo, Prof. Pritesh Purohit	International Research Journal of Engineering and Technology (IRJET), Vol: 04, April 2017	A Study of Some Data Mining Classification Techniques	Teknik klasifikasi tools WEKA untuk *model classification*dengan teknik model untuk Comprehensive set of data *pre-porcessing tool, Environment for comparing learning algorithms.*

Tabel II.2 Penelitian Terkait dengan Algoritma Naïve Bayes dan C4.5

No	Nama Penulis	Publish	Judul	Pembahasan
1	Yuris Alkhalifi, Ainun Zumarniansyah, Rian Ardianto, Nila Hardi, Annisa Elfina Augustia	Jurnal TECHNO Nusa Mandiri Vol.17, No. 1 March 2020	Comparison Of Naive Bayes Algorithm And C.45 Algorithm In Classification Of Poor Communities Receiving Non Cash Food Assistance In Wanasari Village Karawang Regency	Perbandingan 2 algoritma akan dilakukan yaitu Naive Bayes Classifier dan Decision Tree C.45. Total sampel yang digunakan adalah sebanyak 200 kepala data rumah tangga yang kemudian akan dibagi menjadi 2 bagian menjadi teknik validasi yaitu 90% data pelatihan dan 10% data uji dari total sampel yang digunakan maka model yang diusulkan dibuat dalam aplikasi RapidMiner dan kemudian dievaluasi menggunakan tabel Confusion Matrix untuk mengetahui tingkat akurasi tertinggi dari 2 metode ini. Hasil dalam klasifikasi ini menunjukkan bahwa tingkat akurasi dalam metode Naive Bayes Classifier adalah 98,89% dan untuk tingkat akurasi dalam metode Pohon Keputusan C.45 adalah 95,00%. Maka kesimpulan bahwa dalam penelitian ini algoritma dengan tingkat akurasi tertinggi adalah metode algoritma Naive Bayes Classifier dengan perbedaan tingkat akurasi 3,89%
2	Robbi Rahim, Ilka Zufria, Nuning Kurniasih, Muhammad Yasin Simargolang, Abdurrozzaq Hasibuan, Dian Utami Sutiksno, Ricardo Freedom Nanuru, Jusuf Nikolas Anamofa, Ansari Saleh Ahmar, Achmad Daengs GS	International Journal of Engineering & Technology, 7 (2.3) (2018) 68-72	C4.5 Classification Data Mining for Inventory Control	Metode data mining yang dapat diterapkan seperti metode klasifikasi C4.5. Menurut penelitian ini C4.5 atau pohon keputusan memiliki kelebihan seperti hasil dari analisis pohon yang mudah dimengerti, mudah untuk dibuat, hanya membutuhkan sedikit data eksperimental dibanding algoritma klasisifikasi lainnya, mampu memproses data nominal, model hasil dapat dengan mudah dipahami, menggunakan teknik statistik yang dapat divalidasi, waktu komputasi relatif lebih cepat, akurasi yang dapat cocok untuk klasifikasi inventory. Eksperimen yang dilakukan pada peneltian ini menggunakan 14 nilai atribur seperti : Jumbo, Kyodoyushi, cobra, unigo, jumbo, akio, jumbo, idemitsuu, ultraline, trane, fuso, jot, idemitsu dan kayaba. Hasil klasifikasi nilai gain tertinggi adalah case dengan brand = Jumbo dengan tingkat konsumti Tinggi = 1.

No	Nama Penulis	Publish	Judul	Pembahasan
3	Cindy Paramitha Lubis, Zakarias Situmorang	CSRID Journal, Vol. 12 No. 1 Februari 2020	Penerapan Metode Naïve Bayes dan C4.5 Pada Penerimaan Pegawai di Universitas Potensi Utama	Metode Naïve Bayes dan C4.5 tersebut merupakan metode klasifikasi yang diterapkan pada data mining. Tujuan dibuatnya penelitian ini untuk menentukan tingkat akurasi antara kedua metode tersebut berdasarkan ketepatan perhitungan Correctly Classified Instance dan Incorrectly Classified Instance. Pengujian metode pada penelitian ini dilakukan dengan menggunakan *tools* WEKA3.8. Hasil yang didapat Pada metode Naïve Bayes tingkat akurasi yang didapat 77,7778% dan C4.5 memiliki tingkat akurasi 94,444% dari 36 data latih berhasil diuji. Sehingga hasil yang didapat C4.5 merupakan metode yang lebih tepat di gunakan dari pada Naïve Bayes.
4	Pradip Kumar Bala	Proceedings of The 2010 IEEE IEEM	Decision Tree Based Demand Forecasts for Improving Inventory Performance	Membantu manager inventaris untuk mengelola kinerja rantai pasokan mereka dengan lebih baik dengan mengurangi hari kerja dan tingkat layanan secara bersamaan. Mencapai hari sebagai ukuran tingkat persediaan umumnya berhasil dikurangi oleh pengecer pada tingkat biaya layanan di sebagian besar tempat. sebagian besar supermarket besar mengambil jalan ke pohon keputusan untuk berbagai kegunaan dalam CRM meningkatkan manajemen persediaan. Atribut yang digunakan pada penelitian ini adalah penghasilan, rumah, mobil, jumlah anak, usia, domisili, provinsi, jumlah penghasilan anggota. Presentase keseluruhan terhadap nilai ambang batas kepercayaan sebesar 75% seperti yang digunakan pada dunia bisnis.
5	Wisti Dwi Septiani	Jurnal Pilar Nusa Mandiri, Vol. 13 No. 1, Maret 2017	Komparasi Metdoe Klasifikasi Data Mining Algoritma C4.5 dan Naïve Bayes untuk Prediksi Penyakit Hepatitis	klasifikasi data mining metode C4.5 menghasilkan akurasi 77,29% dan nilai AUC 0,846 yang termasuk dalam Good Classification. Naive Bayes menghasilkan akurasi 83,71% dan nilai AUC 0,812. Dengan demikian dapat disimpulkan bahwa kedua metode ini akurat dalam melakukan prediksi untuk penyakit hepatitis. Melihat dari hasil perbandingan kedua algoritma tersebut memang dapat dinyatakan bahwa Algoritma C4.5 lebih unggul dari Naive Bayes karena memiliki nilai AUC 0,846 dengan kategori Good Clasification.

No	Nama Penulis	Publish	Judul	Pembahasan
6	Sayali D. Jadhav, H. P. Channe	International Journal Of Science and Research (IJSR), 2016	Comparative Study of K-NN, Naive Bayes and Decision Tree Classification Techniques	Dari survei dan analisis perbandingan antara algoritma klasifikasi penambangan data (Pohon keputusan, KNN, Bayesian), itu menunjukkan bahwa semua algoritma Pohon Keputusan lebih akurat dan mereka memiliki tingkat kesalahan yang lebih sedikit dan lebih mudah. Algoritma dibandingkan dengan K-NN dan Bayesian. Pengetahuan dalam pohon keputusan direpresentasikan dalam bentuk aturan [JIKA-MAKA] yang lebih mudah dipahami manusia. Pada penilitian ini menggunakan atribut nominal cuaca, tantangan segmen, supermarket, nominal cuaca, segmen medium, pasar super. Dengan nilai node yang tertinggi adalah nominal. Hasil implementasi WEKA pada dataset yang sama menunjukkan bahwa *decision tree* merupakan yang paling unggul
7	Pikulkaew Tangtisanon	International Conference on Computer and Communication Systems, 2018	Web Service Based Food Additive Inventory Management with Forecasting System	Untuk dapat bertahan dalam dunia yang kompetitif, industri harus menemukan solusi manajemen stok yang praktis karena kekurangan stok menyebabkan industri kehilangan kesempatan untuk menjual sementara kelebihan stok menyebabkan defisit. Makalah ini berfokus pada manajemen persediaan dan sistem perkiraan stok. Layanan web diimplementasikan sebagai pendekatan baru untuk sistem manajemen persediaan yang membantu mengelola dan menemukan bahan tambahan makanan yang ada dalam basis data bahan tambahan makanan internasional yang disahkan oleh Codex Alimentarius Commission. MSE Untuk setiap Model Dalam Average Decsion Tree 30.68 Regresi linier 28.26 Naive Bayes 54.35 Support Vektor Machine 82.84
8	Lawana Masankaran, Asst.Prof.Dr.Waraporn Viyanon, Asst.Prof.Dr.Visan Mahasittiwat	IEEE, 2018	Classification of Benign Paroxysmal Positioning Vertigo Types from Dizziness Handicap Inventory using Machine Learning Techniques.	Algoritma terbaik untuk membangun model untuk dataset ini adalah Gaussian Naïve Bayes. Model yang dihasilkan oleh teknik pembelajaran mesin menggunakan dataset DHI ini tidak memiliki efisiensi tinggi untuk digunakan secara kuat untuk membedakan tipe BPPV antara PC-BPPV dan HC-BPPV di rumah sakit. Para dokter masih perlu menggunakan PC-BPPV dan HC-BPPV di rumah sakit. Para dokter masih perlu menggunakan Dix-Hallpike. Akan tetapi, ini dapat menjadi alat untuk memprediksi jenis BPPV sebelum bertemu dengan dokter. Lebih banyak data dapat menghasilkan model kinerja yang lebih baik dan mengurangi masalah overfitting.

No	Nama Penulis	Publish	Judul	Pembahasan
9	Riri Nada Devita, Heru Wahyu Herwanto, Aji Prasetya Wibawa	JTIIK, Vol. 5, No. 4 September 2018	Perbandingan Kinerja Metode Naïve Bayes dan K-Nearest Neighbor Untuk Klasifikasi Artikel Berbahasa Indonesia	Metode Naive Bayes dipilih karena dapat menghasilkan akurasi yang maksimal dengan data latih yang sedikit. Sedangkan metode K-Nearest Neighbor dipilih karena metode tersebut tangguh terhadap data noise. Kinerja dari kedua metode tersebut akan dibandingkan, sehingga dapat diketahui metode mana yang lebih baik dalam melakukan klasifikasi dokumen. Hasil yang didapatkan menunjukkan metode Naive Bayes memiliki kinerja yang lebih baik dengan tingkat akurasi 70%, sedangkan metode K-Nearest Neighbor memiliki tingkat akurasi yang cukup rendah yaitu 40%..
10	Guo-zhi ZHAO and Xiao-bing LIU	International Conference on Advances in Management Science and Engineering (AMSE 218)	Inventory Classification Management Strategy of Parts and Components Within Equipment MRO Enterprises	Strategi inventaris yang tepat untuk komponen dan komponen penting untuk mengendalikan inventaris yang wajar selama proses layanan pemeliharaan di dalam perusahaan MRO. Hasil klasifikasi dari pohon keputusan adalah kebijakan meminimalkan biaya perawatan, yaitu, manajemen persediaan suku cadang dan biaya kekurangan, kualitas pemeliharaan, tingkat penggunaan, biaya persediaan, biaya pembelian, kemampuan peramalan permintaan dan karakter pasokan suku cadang adalah faktor-faktor yang dipertimbangkan dalam model ini. Tingkat *output* Node 5 simpul 3 adalah 1,2,3. Jawaban dari pilihan simpul adalah kategori yan gsesuai dengan nilai maksimumnya.
11	Mücahid Mustafa Saritas , Ali Yasar	International Journal of Intelligent System and Applications in Engineering (IJISAE), 2019	Performance Analysis of ANN and Naive Bayes Classification Algorithm for Data Classification	Klasifikasi adalah teknik penambangan data yang penting dengan berbagai aplikasi untuk mengklasifikasikan berbagai jenis data yang ada di hampir semua bidang kehidupan kita. Kinerja algoritma klasifikasi umumnya diperiksa dengan mengukur keakuratan klasifikasi. Dalam penelitian ini, terlihat bahwa jaringan syaraf tiruan dan algoritma Naïve Bayes dapat digunakan dan hasil yang baik dapat diperoleh dari algoritma klasifikasi. Dalam penelitian kami, diamati bahwa kanker payudara, yang merupakan jenis kanker yang sangat penting, diklasifikasikan dengan akurasi 86,95 dengan JST dan 83,54 dengan algoritma Naïve Bayes menggunakan parameter yang didapat dan dikendalikan secara rutin dari pasien, sehingga menunjukkan bahwa algoritma dapat digunakan untuk deteksi dini kanker payudara.

No	Nama Penulis	Publish	Judul	Pembahasan
12	Pradip Kumar Bala	International Journal of Computer Applications, Vol 10, No.9, November 2010	Purchase-driven Classification for Improved Forecasting in Spare Parts Inventory Replenishment	Sebagian besar praktik manajemen inventori mengambil jalan ke pohon keputusan untuk berbagai penggunaan dalam CRM Pengurangan mencapai hari menyiratkan pengurangan tingkat persediaan dan pengurangan kegagalan penjualan menunjukkan peningkatan tingkat layanan. pohon keputusan untuk berbagai penggunaan dalam CRM,, meningkatkan manajemen inventaris dengan menggunakan aturan pohon keputusan akan menjadi aplikasi tambahan untuk meningkatkan profitabilitas operasi.
13	Gaurav L. Agrawal, Prof. Hitesh Gupta	IJETAE, Vo.3, Issue 3, Maret 2013	Optimization of C4.5 Decision Tree Algorithm for Data Mining Application	Klasifikasi adalah suatu bentuk analisis data yang mengekstraksi model yang menggambarkan kelas data penting. C4.5 adalah salah satu algoritma klasifikasi paling klasik dalam penambangan data C4.5 mengadopsi pendekatan greedy (yaitu, tidak dapat dilacak kembali) di mana pohon keputusan dibangun dalam posisi teratas. Sebagian besar algoritma untuk induksi pohon keputusan juga mengikuti pendekatan top-down.
14	Saruni Dwiasnati and Yudo Devianto	IOP Publishing, 2018	Utilization of Prediction Data for Prospective Decision Customers Insurance Using the Classification Method of C.45 and Naive Bayes Algorithms	Pemanfaatan Data Prediksi untuk Pelanggan Asuransi Prospektif menggunakan Metode Klasifikasi C4.5 dan Naive Bayes dapat disimpulkan sebagai berikut: Algoritma Naive Bayes memiliki tingkat akurasi tertinggi 90,11% sedangkan C4.5 adalah 75,27%, perbedaan di antara mereka adalah 15%. Model algoritma Naive Bayes memiliki AUC 0,961 dan C4.5 0,775, dari nilai AUC, algoritma Naive Bayes termasuk dalam kategori klasifikasi sangat baik dan klasifikasi adil C4.5, sehingga algoritma Naive Bayes dapat diimplementasikan dalam menentukan calon pelanggan asuransi
15	Burhan Rashid Hussein, Asem Kasem, Saiful Omar, and Nor Zainah Siau	Springer Nature Switzerland AG 2019	A Data Mining Approach for Inventory Forecasting: A Case Study of a Medical Store	Salah satu faktor yang sering mengakibatkan kekurangan atau kedaluwarsa pengobatan yang tidak terduga adalah tidak adanya, atau terus menggunakan mekanisme peramalan inventaris yang tidak efektif. Kekurangan yang tak terduga dari obat yang mungkin menyelamatkan jiwa berpotensi menyebabkan kematian, sementara kelebihan persediaan dapat memengaruhi baik anggaran medis maupun penyediaan layanan kesehatan. Atribut yang digunakan pada penetian ini Product_1359, Product_1286, dan Product 0979 sub atributnya adalah max, min, median, minggu, dan varian. Hasil nilai MLP pada Product_1359 sebesar 829290, Product_1286 sebesar 138438, dan Product 0979 sebesar 12801.

Berdasarkan penelitian yang terdahulu untuk perbandingan algoritma antara algoritma C4.5 dan algoritma Naive bayes. Beberapa penelitian ditemukan C4.5 adalah algoritma yang terbaik seperti penelitian yang dilakukan oleh Cindy Paramitha Lubis, Rika Rosnelly, Roslina, Zakarias Situmorang dan penelitian Wisti Dwi Septiani beberapa penelitian juga ada yang menemukan bahwasannya Naive Bayes adalah algoritma yang terbaik seperti penelitian yang dilakukan oleh Yuris Alkhalifi, Ainun Zumarniansyah, Rian Ardianto, Nila Hardi, Annisa Elfina Augustia dan penelitian Saruni Dwiasnati and Yudo Devianto. Oleh sebab itu, pada peneliti ini menganggap perlunya perbandingan algoritma untuk Naive Bayes dan C4.5 untuk pengadaan barang inventory di Universitas Potensi Utama untuk memastikan algoritma mana yang lebih baik untuk mengklasifikasian barang inventory.

II.8 Kontribusi Keilmuan

Pada penelitian ini dapat mengembangkan algoritma Naïve Bayes dan C4.5 pada penelitian pengadaan barang inventory di Universitas Potensi Utama dengan algoritma yang lain agar lebih baik. Adapun kontribusi peneltian ini dapat dijelaskan sebagai berikut:

1. Pada penelitian terdahulu belum ditemukan untuk perbandingan algoritma Naïve Bayes dan C4.5 untuk pengadaan barang inventory.
2. Memberikan pengembangan keilmuan dalam perbandingan antara 2 algoritma yaitu Naïve bayes dan C4.5.

3. Memberikan pengembangan keilmuan untuk mengklasifikasikan data terutama klasifikasi barang, menentukan atribut yang sesuai dengan kajian penelitian, dan memberikan pengetahuan terhadap pengembangan model.