

## **BAB II**

### **TINJAUAN PUSTAKA**

Beberapa penelitian yang telah ada sebelumnya yang berkaitan dengan komparasi beberapa metode data mining seperti yang dilakukan Arto Haara (2012) pada penelitiannya yang membahas tentang Teknik non-parametrik *K-Nearest Neighbours* (K-NN) semakin banyak digunakan dalam masalah kehutanan, terutama dalam penginderaan jauh. Analisis regresi parametrik memiliki keunggulan terkenal teori statistik di baliknya, sedangkan sifat statistik K-NN kurang dipelajari. Dalam studi ini, membandingkan kinerja relatif K-NN dan regresi linier dalam sebuah eksperimen. Memeriksa efek dari tiga sifat yang berbeda dari data dan masalah: 1) efek meningkatkan non-linearitas dari tugas pemodelan, 2) pengaruh asumsi tentang populasi dan 3) pengaruh keseimbangan data sampel. Untuk mengetahui pengaruh ketiga aspek tersebut digunakan data simulasi dan masalah pemodelan sederhana. K-NN dan regresi linier memberikan hasil yang cukup mirip sehubungan dengan rata-rata RMSE. Dalam kedua kasus tersebut, *dataset* pemodelan yang seimbang memberikan hasil yang lebih baik daripada *dataset* yang tidak seimbang. Ketika hasil diperiksa dalam kelas diameter, hasil K-NN kurang bias dibandingkan regresi hasil model, terutama dengan nilai diameter yang ekstrim. Perbedaan meningkat dengan meningkatnya non-linearitas model dan meningkatkan ketidakseimbangan data. Perbedaan antara metode itu lebih jelas ketika bentuk model yang diasumsikan tidak benar-benar tepat.

Rohmana (2014) pada penelitiannya yang membahas tentang membuat aplikasi Jaringan Syaraf Tiruan dan *Naive Bayes* untuk memprediksi penyakit stroke dan membandingkan tingkat akuratan dari kedua metode yang digunakan. Sebuah aplikasi software MATLAB diciptakan untuk mendeteksi seseorang Suspect stroke. Metode yang baik dalam mesin pembelajaran berdasarkan data *training* adalah Jaringan Syaraf Tiruan dan *Naive Bayes*, variabel data faktor gejala penyakit stroke digunakan sebagai data *training* dalam proses pembelajaran dari sistem yang dibuat menentukan prediksi penyakit stroke. Dari 120 data percobaan yang dilakukan, akan dihitung akurasi hasil kerja sistem yang dibagi menjadi data pelatihan dan data pengujian. Diperoleh persentase hasil kerja sistem yaitu Jaringan

Syaraf Tiruan sebesar 71,11 persen, sedangkan *Naive Bayes* sebesar 80,55 persen. *Naive Bayes* lebih akurat daripada Jaringan Syaraf Tiruan dalam hal pengambilan keputusan data baru namun Jaringan Syaraf Tiruan memiliki teknik yang lebih bagus dibandingkan dengan *Naive Bayes*. Jaringan Syaraf Tiruan mempunyai karakteristik belajar dari data sebelumnya.

Heru Sulistiono (2015) pada penelitiannya yang membahas tentang Registrasi adalah kegiatan pendaftaran yang dilakukan setiap semester. Dalam sebuah institusi pendidikan, administrasi sangatlah penting. Jika sebuah pendidikan lembaga yang bermasalah dalam administrasi tidak akan berguna dan tidak bisa bertahan lama. Pendaftaran yang dilakukan oleh mahasiswa lama setiap semester merupakan cara terbaik bagi institusi pendidikan termasuk Universitas Indraprasta PGRI, seperti untuk bisa menentukan jumlahnya kelas yang akan disiapkan di awal semester perkuliahan. Dari data tersebut diketahui penurunan jumlah mahasiswa yang akan hadir untuk semester berikutnya. Sebab, faktor apapun yang menyebabkan banyak kendala dalam pendaftaran mahasiswa. Tujuan dari penelitian ini adalah untuk membuat klasifikasi masalah bagi mahasiswa yang tidak di registrasi, dalam penelitian dilakukan perbandingan algoritma C4.5, *naïve bayes* dan Neural Network yang diterapkan pada data mahasiswa bermasalah dalam registrasi. Pelajaran ini bertujuan untuk mengukur keakuratan studi komparatif algoritma klasifikasi pada 3 potongan-potongan yang sulit didaftarkan oleh siswa. Hasil pengujian untuk mengukur kinerja ketiga algoritma tersebut menggunakan metode pengujian *Cross-Validation*, *Confusion matrix* dan kurva ROC, diketahui bahwa algoritma *Naive Bayes* memiliki nilai yang tinggi nilai akurasi yaitu 91.57%, disusul metode C4.5 dengan akurasi 91.43% dan Neural Network terendah adalah metode dengan nilai akurasi 89.02%.

Laksana Priyo Abadi (2016) pada penelitiannya yang membahas tentang Meningkatkan kualitas pelayanan kepada mahasiswa dengan cara melakukan penilaian Fakultas merupakan salah satu cara Universitas Indraprasta untuk tetap bersaing dengan kompetitor. Selain itu, data pendukung yang diperlukan sebagai dasar pengambilan keputusan berasal dari bagian dan instansi lain sehingga proses pengambilan keputusan membutuhkan waktu yang lama. Untuk data Analisis,

penelitian ini menggunakan teknik analisis deskriptif dan instrumen yang digunakan untuk analisis menentukan prioritas kebijakan yaitu dengan menggunakan Algoritma C45, *Naive Bayes*, dan Neural Network dengan perangkat lunak WEKA. Penelitian ini diharapkan dapat menghasilkan model yang mampu mendukung pengambilan keputusan dalam hal menentukan dosen dengan kinerja terbaik akan dinyatakan sebagai dosen berprestasi setiap tahun.

Amirul Mukminin (2017) pada penelitiannya yang dimana membahas tentang Penentuan jenis tanah pada kedalaman tertentu untuk kebutuhan perencanaan pembangunan perumahan dilakukan berdasarkan data *Cone Penetration Test*. Tujuan penelitian ini untuk mengkomparasi Algoritma C4.5, *Naive Bayes*, dan Neural Network sehingga ditemukan pemodelan yang terbaik untuk mengklasifikasikan tanah. Hasil dari penelitian ini didapatkan algoritma terbaik yaitu Algoritma C4.5. Algoritma C4.5 dalam klasifikasi dua kelas mencapai akurasi 98,45% dan AUC 0,981. Dalam klasifikasi tiga kelas C4.5 juga mencapai akurasi tertinggi (93,21%), demikian juga pada klasifikasi tujuh kelas (83,40%). Hasil penelitian ini menyimpulkan bahwa Algoritma C 4.5 dapat dijadikan pilihan dalam mengklasifikasi tanah untuk pembangunan perumahan.

Azahari (2018) pada penelitiannya yang membahas tentang Prediksi kelulusan dibutuhkan oleh manajemen perguruan tinggi dalam menentukan kebijakan preventif terkait pencegahan dini kasus *drop out*. Lama masa studi setiap mahasiswa bisa disebabkan dengan berbagai faktor. Dengan menggunakan data mining algoritma *naïve bayes* dan Neural Network dapat dilakukan prediksi kelulusan mahasiswa di STMIK Widya Cipta Dharma (WiCiDa) Samarinda. Atribut yang digunakan yaitu, umur saat masuk kuliah, klasifikasi kota asal Sekolah Menengah Atas, pekerjaan ayah, program studi, kelas, jumlah saudara, dan Indeks Prestasi Kumulatif (IPK). Sampel mahasiswa yang lulus dan *drop-out* pada tahun 2011 sampai 2019 dijadikan sebagai data training dan data testing. Sedangkan angkatan 2015–2018 digunakan sebagai data target yang akan diprediksi masa studinya. Sebanyak 3229 mahasiswa, 1769 sebagai data training, 321 sebagai data testing, dan 1139 sebagai data target. Semua data diambil dari data mahasiswa program strata 1, dan tidak mengikut sertakan data mahasiswa D3 dan alih

jenjang/transfer. Dari data testing diperoleh tingkat akurasi hanya 57,63%. Hasil penelitian menunjukkan banyaknya kelemahan dari hasil prediksi *Naive Bayes* dikarenakan tingkat akurasi kevalidannya tergolong tidak terlalu tinggi. Sedangkan akurasi prediksi Neural Network adalah 72,58%, sehingga metode alternatif inilah yang lebih baik. Proses evaluasi dan analisis dilakukan untuk melihat dimana letak kesalahan dan kebenaran dalam hasil prediksi masa studi.

Hendri Mahmud Nawawi (2019) pada penelitiannya yang dimana membahas tentang Penyakit jantung merupakan salah satu dari jenis penyakit mematikan yang penanganannya harus diatasi sesegera mungkin karena bisa terjadi secara mendadak kepada penderitanya. Faktor-faktor penyakit jantung yang dikenali berdasarkan kondisi tubuh seorang penderitanya perlu di ketahui sejak dini sehingga resiko kemungkinan terjadi serangan seketika bisa diminimalisir atau bisa diatasi dengan berbagai cara misalnya pola hidup sehat dan olahraga teratur yang bisa mengatur kesehatan jantung dalam tubuh. Dengan melihat kondisi tubuh seseorang berdasarkan jenis kelamin, tensi darah, usia, apakah perokok atau tidak dan beberapa indikator yang menjadi penyebab seseorang terkena penyakit jantung diuraikan dalam sebuah penelitian dengan menggunakan algoritma Neural Network dan *Naive Bayes* dengan tujuan untuk membandingkan tingkat akurasi terhadap atribut berpengaruh untuk memprediksi penyakit jantung, sehingga hasil dari penelitian ini bisa dijadikan referensi untuk memprediksi apakah seseorang terkena penyakit jantung atau tidak berdasarkan data kesehatan yang dijadikan tolak ukur penelitian.

Mega Luna Sulistia (2020) pada penelitiannya yang membahas tentang Klasifikasi adalah proses pengelompokan data berdasarkan variabel yang diamati untuk memprediksi data baru yang belum diketahui kelasnya. Beberapa metode klasifikasi antara lain *Naive Bayes*, *K-Nearest Neighbor*, dan Neural Network. *Naive Bayes* bertujuan untuk klasifikasi berdasarkan nilai probabilitas dari sifat yang telah ada. *K-Nearest Neighbor* bertujuan untuk klasifikasi berdasarkan sifat tetangga terdekatnya, dimana jumlah tetangga =k, sedangkan Neural Network berdasarkan perilaku jaringan syaraf manusia. Penelitian ini akan membandingkan ketiga metode tersebut dalam klasifikasi Seat Load Factor yang merupakan

persentase keterisian pesawat, dan juga ukuran dalam menentukan nilai keuntungan pada penerbangan suatu maskapai. Faktor yang mempengaruhi yaitu jumlah penumpang, harga tiket, rute, dan waktu penerbangan. Berdasarkan analisis yang dilakukan dengan 47 data diketahui bahwa sistem dengan metode *Naïve Bayes* melakukan kesalahan klasifikasi dalam 14 data sehingga tingkat akurasinya sebesar 70%. Sistem dalam metode *K-Nearest Neighbor* dengan  $k=5$  melakukan kesalahan klasifikasi dalam 5 data, sehingga tingkat akurasinya sebesar 89%, dan sistem Neural Network melakukan kesalahan klasifikasi dalam 10 data sehingga tingkat.

Parastoo Golpour (2020) pada penelitiannya yang dimana membahas tentang Angiografi koroner dianggap sebagai metode yang paling handal untuk diagnosis penyakit kardiovaskular. Namun, angiografi adalah prosedur invasif yang memiliki risiko komplikasi. Oleh karena itu, akan lebih disukai jika metode yang tepat diterapkan untuk menentukan kebutuhan untuk angiografi. Tujuan dari penelitian ini adalah untuk membandingkan mesin vektor pendukung, *naïve bayes*, dan regresi logistik untuk menentukan faktor diagnostik yang dapat memprediksi kebutuhan angiografi koroner. Model ini adalah algoritma pembelajaran mesin. Pembelajaran mesin dipertimbangkan menjadi cabang kecerdasan buatan. Tujuannya adalah untuk merancang dan mengembangkan algoritma yang memungkinkan komputer untuk meningkatkan kinerja mereka dalam analisis data dan pengambilan keputusan. Prosesnya melibatkan analisis pengalaman masa lalu untuk menemukan keteraturan dan pola yang praktis dan bermanfaat, yang mungkin juga diabaikan oleh manusia. Bahan dan Metode: Studi *cross-sectional* ini dilakukan pada 1187 calon angiografi dirujuk ke Rumah Sakit Ghaem, Mashhad, Iran dari 2011 hingga 2012. Sebuah regresi logistik, *naïve bayes*, dan mesin vektor dukungan diterapkan untuk menentukan apakah mereka bisa memprediksi hasil angiografi. Setelah itu, sensitivitas, spesifisitas, positif dan nilai prediksi negatif, AUC (area di bawah kurva), dan akurasi dari ketiga model tersebut dihitung untuk membandingkannya. Semua analisis dilakukan dengan menggunakan perangkat lunak R 3.4.3 (R Core Tim; Auckland, Selandia Baru) dengan bantuan paket perangkat lunak lain termasuk pengoperasian receiver karakteristik (ROC), tanda sisipan, *e1071*, dan *penambang*. Hasil: Area di bawah kurva untuk logistik regresi,

naïve bayes, dan mesin vektor dukungan adalah serupa masing-masing 0.76, 0.74, dan 0.75. Jadi, dari segi kesederhanaan model dan kesederhanaan aplikasinya, maka model naïve bayes dengan tiga variabel memiliki kinerja terbaik dibandingkan dengan model regresi logistik dengan tujuh variabel dan mesin vektor pendukung dengan enam variabel. Kesimpulan: Jenis kelamin, usia, dan glukosa darah puasa (FBG) merupakan faktor terpenting untuk memprediksi hasil penyakit jantung koroner. angiografi. Model *naïve Bayes* berkinerja baik dengan menggunakan tiga variabel ini saja, dan mereka dianggap sebagai variabel penting untuk dua model lainnya juga. Menurut yang bisa diterima prediksi model, mereka dapat digunakan sebagai metode pragmatis, hemat biaya, dan berharga itu mendukung dokter dalam pengambilan keputusan.

Rani Puspita (2020) pada penelitiannya yang dimana membahas tentang BPJS sangat membantu dalam hal kesehatan. Namun ketika ada banyak orang yang menggunakan layanan tersebut, maka akan banyak pula pro dan kontra yang didapatkan. Oleh karena itu peneliti melakukan analisis sentimen dalam bidang data mining terhadap pengguna BPJS pada media sosial Twitter sebanyak 1000 data yang kemudian di filtering menjadi 903 data dikarenakan adanya data yang terduplikat. Penulis menggunakan metode KNN, *Decision Tree*, dan *Naïve Bayes* untuk memperbandingkan tingkat akurasi dari ketiga metode tersebut. Peneliti menggunakan *tools RapidMiner* versi 9.7.2. Hasil penelitian menunjukkan bahwa analisis sentimen terhadap data *Twitter* terhadap layanan BPJS dengan menggunakan metode KNN mencapai tingkat akurasi 95.58% dengan *class precision* untuk pred. *negative* adalah 45.00%, pred. *positive* adalah 0.00%, dan pred. *neutral* adalah 96.83%. Lalu pada metode *Decision Tree* tingkat akurasinya mencapai 96.13% dengan *class precision* untuk pred. *negative* adalah 55.00%, pred. *positive* adalah 0.00%, dan pred. *neutral* adalah 97.28%. Dan yang terakhir adalah metode *Naïve Bayes* yang mencapai akurasi 89.14% dengan *class precision* untuk pred. *negative* adalah 16.67%, pred. *positive* adalah 1.64%, dan pred. *neutral* adalah 98.40%.

Tyas Yanotama (2020) dalam penelitiannya yang membahas tentang Universitas Teknologi Digital Indonesia Setiap tahunnya melakukan penerimaan

mahasiswa baru yang dilakukan oleh bagian marketing dan admisi, penerimaan mahasiswa baru sangat penting untuk Universitas Teknologi Digital Indonesia Karena operasional kampus dibiayai oleh pemasukan yang berasal dari SPP mahasiswa Universitas Teknologi Digital Indonesia Mengharapkan jumlah mahasiswa baru setiap tahunnya sesuai target, sehingga bagian marketing dan admisi Universitas Teknologi Digital Indonesia Memiliki jumlah target penerimaan mahasiswa baru setiap tahunnya, perlunya target penerimaan mahasiswa baru digunakan untuk memberikan gambaran jumlah mahasiswa baru di tahun selanjutnya, atribut yang digunakan dalam penelitian antara lain gelombang pendaftaran, jurusan, pendapatan orang tua, biaya kuliah dan nilai rata-rata UN. *Naïve Bayes* digunakan untuk memprediksi jumlah mahasiswa baru dengan menggunakan data pendaftar ulang di tahun sebelumnya yang memiliki atribut yaitu asal kota, gelombang, program studi, penghasilan orang tua, nilai UN dan status registrasi, sehingga pihak marketing dan admisi Universitas Teknologi Digital Indonesia Mendapat gambaran jumlah mahasiswa baru di tahun depan Hasil dari penelitian ini adalah sistem yang dapat memprediksi data dengan kelas yaitu registrasi dan tidak registrasi Dari 1704 data testing yang diproses menggunakan sistem didapatkan hasil prediksi registrasi sebanyak 1226 data dan tidak registrasi 478 data Untuk pengujian dari 731 data didapatkan hasil prediksi 679 data terprediksi benar dan 52 data salah prediksi Tingkat akurasi probabilitas yang didapatkan sebesar 92,88%

Kaitan dari semua penelitian yang telah ada sebelumnya yang berkaitan dengan komparasi beberapa metode data mining dengan penelitian yang akan dilaksanakan adalah untuk mencari ketepatan atau akurasi dari pada beberapa metode data mining tersebut. Untuk lebih jelas daripada inti penelitian sebelumnya telah terangkum dalam tabel 2.1.

Tabel 2.1 Tinjauan Pustaka

No	Nama, Tahun	Permasalahan	Data	Hasil
1.	Arto Haara, 2012	Membandingkan kinerja relatif K-NN dan Regresi Linier dalam sebuah eksperimen dengan menggunakan pemodelan sederhana	Dataset kehutanan (Pohon) yang seimbang dan tidak seimbang	Memberikan kesimpulan bahwa K-NN dan regresi linier memberikan hasil yang cukup mirip. Dalam kedua kasus tersebut, Dataset pemodelan berimbang memberikan hasil yang lebih baik daripada set data tidak seimbang.
2.	Rohmana, 2014	Membandingkan Jaringan Syaraf Tiruan dan <i>Naive Bayes</i> untuk memprediksi penyakit stroke dan membandingkan tingkat akuratan dari kedua metode yang digunakan.	120 Data faktor gejala penyakit stroke	Informasi bahwa <i>Naive Bayes</i> lebih akurat daripada Jaringan Syaraf Tiruan dalam hal pengambilan keputusan data baru namun Jaringan Syaraf Tiruan memiliki teknik yang lebih bagus dibandingkan dengan <i>Naive Bayes</i> . Jaringan Syaraf Tiruan mempunyai karakteristik belajar dari data sebelumnya.
3.	Heru Sulistiono, 2015	Membandingkan Algoritma c4.5, Neural Network	747 Data mahasiswa Universitas	Memberikan informasi tentang algoritma <i>Naive Bayes</i> merupakan

No	Nama, Tahun	Permasalahan	Data	Hasil
		dan <i>Naïve Bayes</i> untuk klasifikasi mahasiswa yang bermasalah dalam registrasi	Indraprasta PGRI	algoritma terbaik dan dapat memberikan pemecahan dalam permasalahan mahasiswa yang bermasalah dalam registrasi.
4.	Laksana Priyo Abadi, 2016	Membandingkan Algoritma C4.5, <i>Naive Bayes</i> , Dan Neural Network dalam pemilihan dosen teladan di UNIVERSITAS INDRAPRASTA	75 Data dosen UNIVERSITAS INDRAPRASTA	Memberikan informasi tentang algoritma <i>Naïve Bayes</i> merupakan algoritma terbaik dan dapat memberikan pemecahan dalam permasalahan pemilihan dosen teladan UNIVERSITAS INDRAPRASTA
5.	Amirul Mukminin. 2017	Perbandingan Algoritma C4.5, <i>Naïve Bayes</i> Dan Neural Network Untuk Penentuan jenis tanah pada kedalaman tertentu untuk kebutuhan perencanaan pembangunan perumahan	Data klasifikasi tanah	Memberikan informasi tentang algoritma terbaik yaitu Algoritma C4.5. Algoritma C4.5 dalam klasifikasi dua kelas mencapai akurasi 98,45%. Dalam klasifikasi tiga kelas C4.5 juga mencapai akurasi tertinggi (93,21%), demikian juga pada klasifikasi tujuh kelas (83,40%). Hasil

No	Nama, Tahun	Permasalahan	Data	Hasil
				<p>penelitian ini menyimpulkan bahwa Algoritma C 4.5 dapat dijadikan pilihan dalam mengklasifikasi tanah untuk pembangunan perumahan.</p>
6.	Azahari, 2018	<p>Membandingkan <i>Naive Bayes</i> Dan Neural Network untuk memprediksi Masa Studi Mahasiswa S1 STMIK Widya Cipta Dharma (WiCiDa) Samarinda</p>	<p>3229 Data Mahasiswa STMIK Widya Cipta Dharma Samarinda</p>	<p>Memberikan informasi bahwa banyaknya kelemahan dari hasil prediksi <i>Naive Bayes</i> dikarenakan tingkat akurasi kevalidannya tergolong tidak terlalu tinggi. Sedangkan akurasi prediksi Neural Network adalah 72,58%, sehingga metode alternatif inilah yang lebih baik.</p>
7.	Hendri Mahmud Nawawi, 2019	<p>Membandingkan tingkat akurasi Neural Network Dan <i>Naive Bayes</i> dengan atribut yang berpengaruh untuk memprediksi penyakit jantung</p>	<p>Data factor gejala penyakit jantung dengan 14 atribut</p>	<p>Memberikan informasi bahwa kedua metode tersebut dapat memprediksi penyakit jantung dengan nilai 75%, sehingga memiliki akurasi yang baik, namun Algoritma Neural Network dan Algoritma <i>Naive Bayes</i> memiliki</p>

No	Nama, Tahun	Permasalahan	Data	Hasil
				nilai klasifikasi AUC yang buruk dengan nilai 50%
8.	Mega Luna Sulistia, 2020	Membandingkan <i>Naive Bayes</i> , <i>K-Nearest Neighbor</i> , dan Neural Network <i>Classification</i> untuk klasifikasi <i>Seat Load Factor</i> yang merupakan persentase keterisian pesawat, dan juga ukuran dalam menentukan nilai keuntungan pada penerbangan suatu maskapai.	47 Data penumpang penerbangan <i>Outbound</i> Lombok	Memberikan informasi bahwa <i>Naive Bayes</i> melakukan kesalahan klasifikasi dalam 14 data sehingga tingkat akurasi sebesar 70%. Sistem dalam metode <i>K-Nearest Neighbor</i> dengan $k=5$ melakukan kesalahan klasifikasi dalam 5 data, sehingga tingkat akurasi sebesar 89%, dan sistem Neural Network melakukan kesalahan klasifikasi dalam 10 data sehingga tingkat akurasi sebesar 78%. Metode yang memiliki nilai akurasi tertinggi merupakan metode terbaik yang akan digunakan
9.	Parastoo Golpour, 2020	Memilih metode terbaik antara <i>Support Vector</i>	1187 Data pasien jantung koroner dari tahun 2011-	Memberikan informasi bahwa Jenis kelamin, usia, dan puasa Menilai

No	Nama, Tahun	Permasalahan	Data	Hasil
		<p><i>Machine, Naïve Bayes</i> dan <i>Logistic Regression</i> untuk menilai Kebutuhan Angiografi Koroner</p>	<p>2012 pada rumah sakit Ghaem, Iran</p>	<p>kebutuhan glukosa darah (FBG) ditemukan menjadi faktor paling penting untuk memprediksi hasil koroner. angiografi. Model <i>naïve Bayes</i> berkinerja baik dengan menggunakan tiga variabel ini saja, dan mereka dianggap sebagai variabel penting untuk dua model lainnya juga. Menurut yang bisa diterima prediksi model, mereka dapat digunakan sebagai metode pragmatis, hemat biaya, dan berharga itu mendukung dokter dalam pengambilan keputusan.</p>
10.	Rani Puspita, 2020	<p>Membandingkan Metode KNN, <i>Decision Tree</i>, dan <i>Naïve Bayes</i> terhadap analisis sentimen pengguna layanan BPJS pada media social twitter.</p>	<p>1000 Data <i>Twitter</i> terhadap layanan BPJS</p>	<p>Memberikan informasi dengan menggunakan metode KNN mencapai tingkat akurasi 95.58% dengan <i>class precision</i> untuk pred. <i>negative</i> adalah 45.00%, pred. <i>positive</i> adalah 0.00%,</p>

No	Nama, Tahun	Permasalahan	Data	Hasil
				<p>dan pred. <i>neutral</i> adalah 96.83%. Lalu pada metode <i>Decision Tree</i> tingkat akurasinya mencapai 96.13% dengan <i>class precision</i> untuk pred. <i>negative</i> adalah 55.00%, pred. <i>positive</i> adalah 0.00%, dan pred. <i>neutral</i> adalah 97.28%. Dan yang terakhir adalah metode <i>Naïve Bayes</i> yang mencapai akurasi 89.14% dengan <i>class precision</i> untuk pred. <i>negative</i> adalah 16.67%, pred. <i>positive</i> adalah 1.64%, dan pred. <i>neutral</i> adalah 98.40%.</p>
11.	Tyas Yanotama, 2020	Apakah metode <i>naïve bayes</i> dapat memprediksi calon mahasiswa baru Universitas Teknologi Digital Indonesia Yang melakukan Herregistrasi	2899 Data calon mahasiswa baru STMIK AKAKOM Yogyakarta	Memberikan informasi bahwa metode <i>naïve bayes</i> dapat memprediksi calon mahasiswa baru Universitas Teknologi Digital Indonesia Yang melakukan Herregistrasi dengan akurasi 92.88%

Tabel 2.2 Usulan Penelitian

No	Nama, Tahun	Permasalahan	Data	Kontribusi
1.	Ovandry C.J. 2021	Memilih metode terbaik antara Metode <i>Naïve Bayes</i> , <i>KNN</i> , dan <i>Decision Tree</i> dalam hal akurasi untuk memprediksi calon mahasiswa baru Universitas Teknologi Digital Indonesia Yang melakukan Herregistrasi.	1001 Data calon mahasiswa baru Universitas Teknologi Digital Indonesia Dari tahun 2013-2019 yang telah di <i>filtering</i>	Diharapkan memberikan informasi tentang metode terbaik antara Metode KNN, Neural Network, Naïve Bayes dan Decision Tree dalam hal akurasi