

BAB V

ANALISA PROJECT INNOVATION: INNOVATION MANAGEMENT

Proyek magang yang penulis lakukan adalah inovasi manajemen, dengan fokus pada data akademik yang dimiliki Universitas Ma Chung untuk memprediksi kelulusan tepat waktu mahasiswa Program Sarjana (S1) Universitas Ma Chung dengan menggunakan model *Decision Tree* dan *Knime* sebagai alat analisis.

Dalam bab ini penulis akan membahas tinjauan teori dan penelitian yang pernah dilakukan terkait dengan prediksi kelulusan tepat waktu mahasiswa.

5.1. Landasan Teori

1. Data Mining

Pengertian data mining menurut Gartner Group adalah proses menemukan korelasi, pola, dan tren baru yang bermakna dengan menyaring sejumlah besar data yang disimpan dalam repositori, menggunakan teknologi pengenalan pola serta statistik dan teknik matematika. (Larose, 2005)

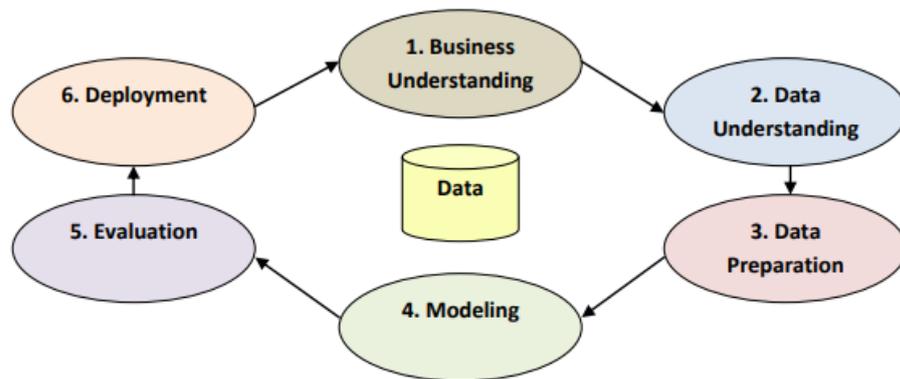
Data Mining merupakan perpaduan dari statistik terapan, logika, kecerdasan buatan, *machine learning* dan sistem manajemen data. (North, 2012)

Proses data mining berdasarkan CRISP-DM (*Cross Industry Standard Process for Data Mining*), terdiri dari 6 fase, yaitu:

- a. *Business/organizational Understanding Phase* (Fase Pemahaman Bisnis)
- b. *Data Understanding Phase* (Fase Pemahaman Data)
- c. *Data Preparation Phase* (Fase Pengolahan Data)
- d. *Modeling Phase* (Fase Pemodelan)

e. *Evaluation Phase* (Fase Evaluasi)

f. *Deployment Phase* (Fase Penyebaran).



Gambar 5. Fase CRISP-DM

Sumber: Data mining for the masses, North (2012)

a. Business/Organizational Understanding Phase

Fase *Business Understanding* adalah merupakan langkah awal melakukan CRISP DM (*Cross Industry Standard Process for Data Mining*). Pada fase ini pengetahuan akan objek bisnis atau organisasi sangat dibutuhkan, bagaimana membangun atau mendapatkan data, mencocokkan tujuan pemodelan sehingga model terbaik dapat dibangun. Kegiatan yang dilakukan antara lain: menentukan tujuan dan persyaratan dengan jelas secara keseluruhan, menerjemahkan tujuan tersebut serta menentukan pembatasan dalam perumusan masalah data mining, dan selanjutnya mempersiapkan strategi awal untuk mencapai tujuan tersebut (Larose, 2006).

b. *Data Understanding Phase* (Fase Pemahaman Data)

Pada fase ini proses mengumpulkan, mengidentifikasi, dan memahami data yang dimiliki organisasi sebelum dilanjutkan dengan tahapan selanjutnya *Data preparation*.

c. *Data Preparation Phase* (Fase Preparasi Data)

Data muncul dalam berbagai bentuk dan format antara lain: numerik, paragraph teks, gambar yang berupa bagan, grafik dan peta, narasi seperti komentar pada survey kepuasan pelanggan atau transkrip kesaksian.

Preparasi data melibatkan sejumlah kegiatan termasuk didalamnya adalah menggabungkan dua atau lebih kumpulan data, mengurangi kumpulan data menjadi hanya variabel data tertentu, membersihkan data dari anomali seperti pengamatan outlier atau data yang hilang, atau memformat ulang data untuk tujuan konsistensi.

d. *Modeling Phase* (Fase Pemodelan)

Model dalam data mining adalah representasi komputerisasi dari pengamatan dunia nyata. Model adalah algoritma untuk mencari, mengidentifikasi, dan menampilkan pola atau pesan apa pun dalam data.

e. *Evaluation Phase* (Fase Evaluasi)

Fase evaluasi digunakan untuk menentukan seberapa berharga model yang diterapkan pada tahap sebelumnya dan apa yang mungkin dilakukan dengan model tersebut.

f. Deployment Phase (Fase Penyebaran)

Aktivitas dalam fase ini termasuk menyiapkan otomatisasi model yang terintegrasi dengan sistem informasi manajemen atau operasional yang ada, memasukkan pembelajaran baru dari penggunaan model kembali ke dalam model untuk meningkatkan akurasi dan kinerja, dan pemantauan serta pengukuran hasil penggunaan model. Dengan mengkomunikasikan fungsi dan kegunaan model secara jelas kepada pemangku kepentingan, secara menyeluruh menguji dan membuktikan model, kemudian merencanakan dan memantau implementasinya,

2. Peran Utama Data Mining

Peran utama data mining menurut Larose (Larose, 2005) adalah:

a. Estimasi

Estimasi merupakan proses untuk menemukan suatu karakteristik penting dari data dalam suatu basis data dengan cara menduga atau menebak dari sebuah nilai yang belum diketahui. Model yang digunakan untuk estimasi diantaranya adalah: *Linier Regresion, Neural Network (NN), Support Vector Machine, Generalized Liniear Model (GLM)*.

b. Forecasting/prediksi

Proses untuk menemukan jenis, pola atau nilai dari data dengan memperkirakan hal tersebut yang belum diketahui untuk masa mendatang. Model yang digunakan untuk forecasting diantaranya adalah: *Linier Regresion, Neural Network (NN), Support Vector Machine, Generalized Liniear Model (GLM)*.

c. Klasifikasi

Klasifikasi merupakan proses menemukan sebuah model atau fungsi yang mendeskripsikan dan membedakan data ke dalam kelas-kelas. Klasifikasi melibatkan proses pemeriksaan karakteristik dari objek dan memasukkan objek ke dalam salah satu kelas yang sudah didefinisikan sebelumnya. Model yang digunakan untuk klasifikasi diantaranya adalah: *Decision tree*, *Naïve Bayes*, *K-Nearest Neighbour(KNN)*, *Logistic Regresion(LoR)*, *Linear Discriminant Analysis (LDA)*.

d. Klustering

Klustering merupakan pengelompokan data tanpa berdasarkan kelas data tertentu ke dalam kelas objek yang sama. Sebuah kluster adalah kumpulan record yang memiliki kemiripan suatu dengan yang lainnya dan memiliki ketidakmiripan dengan record dalam kluster lain. Tujuannya adalah untuk menghasilkan pengelompokan objek yang mirip satu sama lain dalam kelompok-kelompok. Semakin besar kemiripan objek dalam suatu kluster dan semakin besar perbedaan tiap kluster maka kualitas analisisnya semakin baik. Model yang digunakan untuk klustering diantaranya adalah: *K-Means*, *Kmedoids*, *Self Organizing Map (SOM)*, *Fuzzy C-means*.

e. Asosiasi

Asosiasi merupakan proses menemukan atribut yang muncul dalam suatu waktu. Dalam dunia bisnis disebut analisis keranjang belanja (market basket analysis). Tugas asosiasi berusaha untuk mengungkap aturan untuk mengukur hubungan

antara dua atau lebih atribut. Model yang digunakan untuk asosiasi diantaranya adalah: *Coefficient of Correlation, Chi Square, FP-Growth*

3. *Decision Tree*

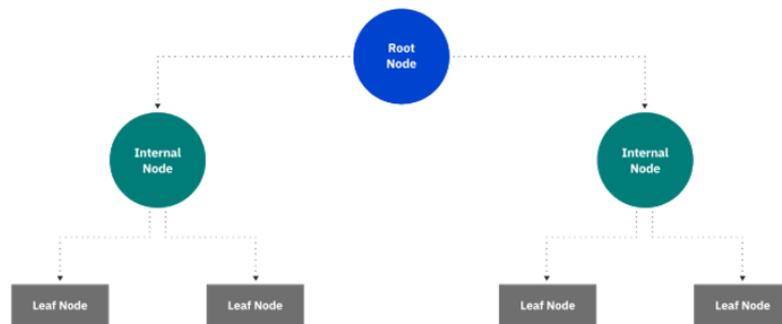
Machine learning merupakan sains (dan seni) memprogram komputer dengan belajar dari data. *Machine learning* memungkinkan sistem membuat keputusan secara mandiri tanpa dukungan dari eksternal dalam bentuk apapun (Theobald, 2017).

Machine Learning terdiri dari *supervised learning*, *unsupervised* dan *reinforcement*. *Supervised learning* merupakan jenis yang populer untuk melakukan operasi *machine learning* dan banyak digunakan untuk data di mana ada pemetaan yang tepat antara data input-output. Kumpulan data diberi label, artinya algoritma mengidentifikasi fitur secara eksplisit dan melakukan prediksi atau klasifikasi yang sesuai. Pada algoritma *unsupervised learning*, data tidak secara eksplisit diberi label ke dalam kelas yang berbeda (tidak memiliki label pada data). Model mampu belajar dari data dengan menemukan pola implisit. Algoritma *unsupervised learning* mengidentifikasi data berdasarkan kepadatan, struktur, segmen serupa, dan fitur serupa lainnya. *Reinforcement learning* berbeda dengan *supervised* maupun *unsupervised learning*. Algoritma ini dimaksudkan untuk membuat komputer dapat belajar sendiri dari lingkungan (*environment*) melalui sebuah agen. Jadi komputer akan melakukan pencarian sendiri (*self discovery*) dengan cara berinteraksi dengan lingkungan.

Decision tree adalah jenis algoritma *supervised learning* yang digunakan untuk masalah klasifikasi dan regresi. Terdapat dua tipe *decision tree* yaitu

classification tree dan *regression tree*. Pada tipe *classification tree*, hasil yang diprediksi bersifat diskrit atau berlawanan contohnya ya atau tidak, kalah atau menang benar atau salah, sedangkan *regression tree* menghasilkan data kontinu atau masih berhubungan. Hasilnya pun bisa dianggap sebagai bilangan nyata, seperti durasi rawat inap pasien hingga harga rumah di suatu daerah.

Sebuah *decision tree* memiliki hirarki seperti halnya struktur pohon, yaitu: *root node* (simpul akar), *branch* (cabang), *internal node* (simpul internal) dan *leaf node* (simpul daun) seperti dalam gambar 6. (<https://www.ibm.com>, n.d.)



Gambar 6. Hirarki Decision Tree

Sumber: (<https://www.ibm.com>, n.d.)

Decision tree adalah *top-down* pohon rekursif dari algoritma induksi, yang menggunakan ukuran seleksi atribut untuk memilih atribut yang diuji. Algoritma *decision tree* mencoba untuk meningkatkan akurasi dengan menghapus cabang-cabang pohon yang mencerminkan noise dalam data. (Han, J, Kamber, M & Pei, 2011)

Jika dibandingkan dengan metode lain beberapa kelebihan *decision tree* dalam menyusun algoritma untuk *machine learning*, di antaranya:

- a. cocok untuk visualisasi data karena mudah dipahami dan diinterpretasikan.

- b. data yang digunakan dalam prosesnya hanya memerlukan persiapan minimal, tidak melalui tahapan *one-hot encoding* (kode biner 0 atau 1)
- c. struktur yang ada bisa ditambahkan opsi baru .
- d. mampu memilih opsi yang terbaik dari seluruh opsi yang tersedia.
- e. Dapat bekerja dengan variabel numerik maupun kategoris.
- f. Pemilihan variabel dilakukan secara otomatis. Itu berarti, jika terdapat variabel yang tidak penting, tidak akan memengaruhi hasil akhir meski sebenarnya ada variabel yang saling bergantung (multikolinearitas).

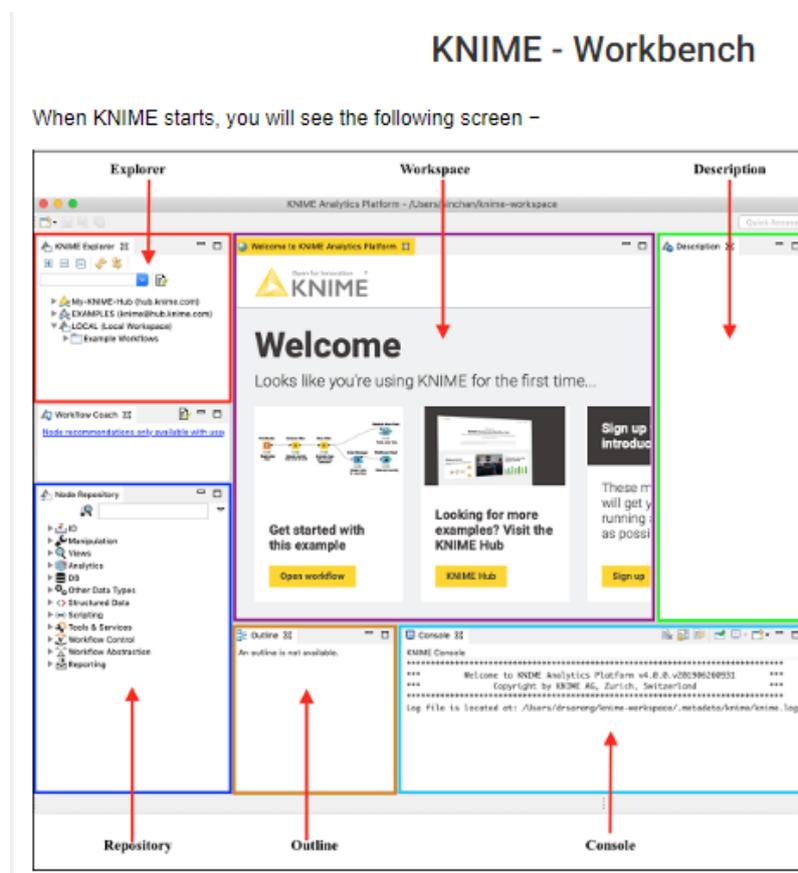
Kelemahan utama *decision tree* terletak pada struktur itu sendiri. Sifat struktur *decision tree* yang sangat terbuka dapat membuatnya menjadi sangat rumit,

4. Knime

KNIME(Konstanz Information Miner) adalah platform data mining untuk analisis, pelaporan dan integrasi data yang termasuk software *opensource* (<https://www.knime.com/software-overview>, n.d.). Knime mulai dikembangkan tahun 2004 oleh tim pengembang perangkat lunak dari Universitas Konstanz, yang dipimpin oleh Michael Berthold, yang awalnya digunakan untuk penelitian di industri farmasi. Mulai banyak digunakan sejak tahun 2006 sehingga tahun 2017 masuk ke *magic quadrant for data science platform* (Gartner Group).

Penulis akan menjelaskan tentang tampilan dan proses analisa menggunakan aplikasi Knime([Www.Tutorialspoint.Com/Knime](http://www.Tutorialspoint.Com/Knime), n.d.) secara garis besar sebagai berikut :

1. ketika menjalankan knime akan tampil di layar *Knime-Workbench*(gambar 7), yang berisikan *Workspace*, *outline*, *repositori node*, *explorer*, *console* dan *description* seperti pada gambar 5.2.



Gambar 7. Workbench Knime

Sumber: ([Www.Tutorialspoint.Com/Knime](http://www.Tutorialspoint.Com/Knime), n.d.)

Workspace merupakan kertas kerja dalam melakukan analisa, *outline* merupakan tampilan miniatur kerangka kerja, *Node* mendefinisikan jenis fungsionalitas tertentu yang digunakan dalam workflow, konsol menyediakan

tampilan berbagai pesan saat menjalankan alur kerja, *explorer* mencantumkan ruangan dimana file *Knime* berada dan *description* menjelaskan item yang sedang dipilih di *workspace*.

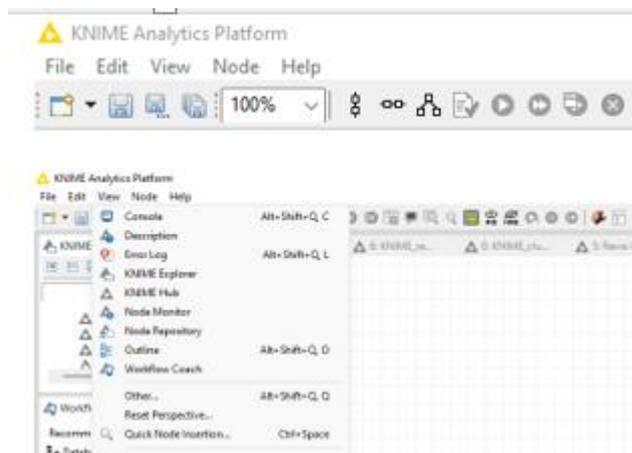
2. Toolbar yang berisikan ikon untuk memfasilitasi tindakan cepat dengan melakukan klik mouse seperti tampak dalam gambar 8.



Gambar 8. Toolbar Knime

Sumber: (Www.Tutorialspoint.Com/Knime, n.d.)

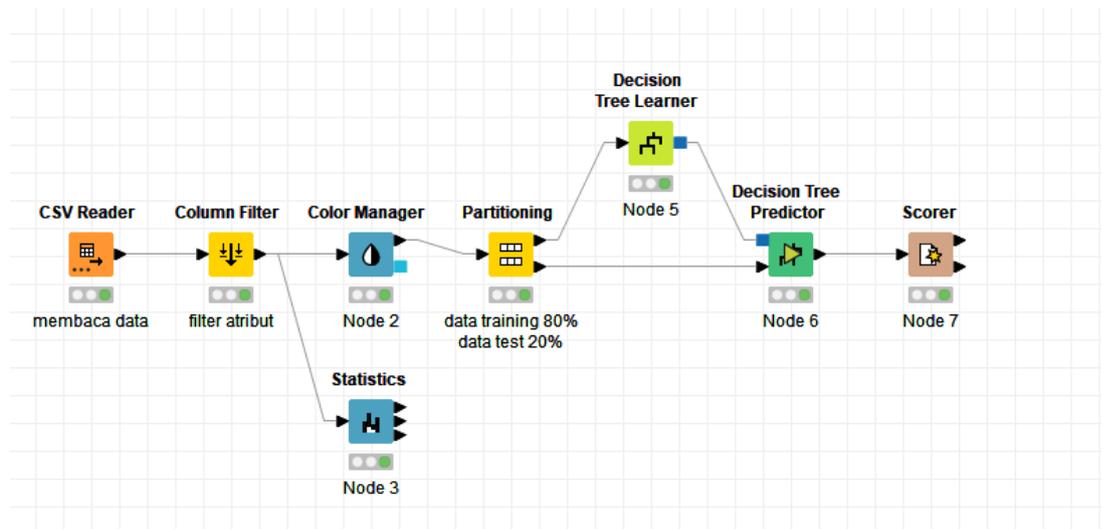
3. Menu yang terdapat pada *Knime* terdiri dari *File*, *Edit*, *View*, *Node* dan *Help* seperti tampak pada gambar 9.



Gambar 9. Menu Knime

Sumber: (Www.Tutorialspoint.Com/Knime, n.d.)

4. Untuk melakukan analisa dengan Knime sangatlah mudah, cukup menggunakan *Node Analytics* yang tersedia, contohnya menentukan berbagai algoritma *machine learning* seperti *Decision Tree*, *Naïve Bayes*, *Clustering*, *Ensemble Learning*, dan sebagainya.



Gambar 10. Node Knime

Sumber: Penulis (2021)

5. Inovasi

Inovasi memiliki fungsi yang khas bagi wirausahawan. Dengan inovasi, wirausahawan menciptakan baik sumber daya produksi baru maupun pengolahan sumber daya yang ada dengan peningkatan nilai potensi untuk menciptakan modal (Drucker, 2012)

Menurut Keeley (Keeley, 2013) terdapat sepuluh tipe inovasi seperti dalam gambar 11, yaitu :

a. Profit Model

Inovasi ini menjelaskan cara baru perusahaan menghasilkan uang, dengan memanfaatkan penawaran perusahaan dan sumber lainnya. Dalam hal ini pemahaman secara mendalam perusahaan terhadap apa yang dilakukan pelanggan akan menimbulkan peluang mendapatkan uang.

b. *Network*

Inovasi ini menjelaskan cara perusahaan memanfaatkan networknya untuk bersama-sama menghasilkan value. Perusahaan akan berkolaborasi dengan perusahaan lain .

c. *Structure*

Perusahaan berinovasi dengan menggunakan semua aset baik yang berwujud ataupun tidak untuk menghasilkan nilai ataupun bisnis yang optimal

d. *Process*

Inovasi yang dilakukan berbasis proses bisnis, mencakup metodologi ataupun kemampuan yang lebih tinggi daripada perusahaan yang sejenis. Perusahaan memiliki keunikan dalam menyampaikan produk ataupun memberikan layanan.

e. *Product Performance*

Inovasi yang dilakukan dengan mengembangkan fitur dan fungsi yang berbeda dari perusahaan lain. Inovasi kinerja produk membahas nilai, fitur, dan kualitas barang yang ditawarkan.

f. *Product System*

Inovasi dalam sistem penciptaan produk, perusahaan membuat pelengkap dari produk dan layanan. Perusahaan berinovasi membuat bundling antara produk dan layanan, untuk menciptakan sistem yang kuat dan terukur. Bundling produk, atau mengambil beberapa produk terkait dan menjualnya dalam satu paket, adalah contoh inovasi Sistem Produk

g. *Service*

Inovasi berbasis layanan, mendukung dan memperkuat nilai produk.

Inovasi layanan memastikan dan meningkatkan utilitas, kinerja, dan nilai dari suatu produk. Produk lebih mudah untuk dicoba, digunakan, dan dinikmati sehingga membuat pelanggan datang kembali.

h. Channel

Inovasi saluran distribusi mencakup semua cara menghubungkan produk perusahaan dengan pelanggan. Saluran *e-commerce* telah muncul sebagai kekuatan dominan tetapi saluran tradisional seperti toko masih penting. Inovasi saluran untuk memastikan bahwa pengguna dapat membeli apa yang mereka inginkan, kapan dan bagaimana mereka menginginkannya, dengan sedikit biaya dan tetap mendapat kepuasan yang maksimal.

i. Brand

Inovasi merek membantu memastikan bahwa pelanggan dan pengguna mengenali, mengingat, dan memilih produk perusahaan daripada pesaing.

j. Customer Engagement

Keakraban dengan pelanggan juga dapat di inovasikan sehingga menghasilkan sebuah *value* yang sangat dihargai oleh pelanggan.



Gambar 11. Ten Types of Innovation

Sumber: (Keeley, 2013)

5.2. Penelitian Terkait

Beberapa penelitian yang menjadi acuan penulis dalam melakukan penelitian proyek inovasi manajemen ini adalah sebagai berikut:

1. Romadhona, dkk melakukan penelitian tentang prediksi kelulusan mahasiswa tepat waktu berdasarkan NIM, usia, jenis kelamin, dan indeks prestasi selama 4 semester pertama (semester 1,2,3,4) menggunakan komparasi algoritma klasifikasi yaitu algoritma *Decision Tree C4.5*, *ID3* dan *Chaid*. Dari ketiga metode yang digunakan tersebut didapat kesimpulan bahwa metode *Decision Tree* menghasilkan kinerja yang lebih baik dengan akurasi 91,51% daripada kedua metode lainnya dan faktor yang paling mempengaruhi kelulusan tepat waktu adalah indeks prestasi semester 4 (Romadhona et al., 2017);
2. Eko Prasetyo Rohmawan melakukan penelitian prediksi kelulusan mahasiswa tepat waktu menggunakan metode *decision tree* dan *artificial neural network*.

Penelitian ini menggunakan data mahasiswa lulus sebagai data training dan data mahasiswa yang belum lulus sebagai data tes. Atribut yang digunakan adalah NIM, Jenis kelamin, asal sekolah, jalur masuk, nilai ujian nasional, gaji orangtua, Indeks Prestasi semester 1-4, Indeks Prestasi Kumulati semester 1-4 dan keterangan lulus. Dari kedua metode tersebut didapat bahwa tingkat akurasi dengan *Artificial Neural Network* yaitu 79,4%, lebih tinggi dari *Decision Tree* yang akurasinya 74,51%. (Rohmawan, 2018);

3. Nurhayati, dkk melakukan penelitian prediksi tingkat kelulusan tepat waktu mahasiswa menggunakan teknik klasifikasi dengan menggunakan metode Naïve Bayes pada Universitas XYZ pada 10 program studi dengan atribut nomer induk, indeks prestasi semester, indeks prestasi kumulatif, jumlah sks dan kelulusan, menunjukkan nilai akurasi 80,95%. (Septianti et al., 2020);
4. Budiyantra dan Irwansyah melakukan penelitian prediksi mahasiswa lulus tepat waktu menggunakan Algoritma Decision Tree (C4.5) dengan atribut NIM, Jenis kelamin, usia, jurusan, kelas, pekerjaan, Indeks Prestasi semester 1 hingga 4 dan keterangan terlambat/tepat. Akurasi yang diperoleh adalah 98,04%.

Dari berbagai penelitian terdahulu mengenai prediksi kelulusan diketahui beberapa kesamaan dalam pemilihan variabel yang menjadi dasar penelitian adalah nomer induk mahasiswa, nilai Indeks prestasi semester 1, 2, 3 dan 4 dan keterangan lulus tepat/ tidak tepat waktu, sedangkan variabel lain lebih bervariasi antara satu penelitian dengan penelitian lainnya berupa : jenis kelamin, usia, indeks prestasi kumulatif dan jumlah sks yang diambil.