

PENERAPAN TEKNIK “ DATA MINING “ DENGAN MENGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER UNTUK MEMPREDIKSI KELULUSAN MAHASISWA

IMPLEMENTATION OF “ DATA MINING “ TECHNIQUES USING NAÏVE BAYES CLASSIFIER ALGORITHM TO PREDICT STUDENT STUDY PERIOD

Ryan Fadillah¹ , Imelda Pangaribuan²

¹Universitas Komputer Indonesia

²Universitas Komputer Indonesia

Email : patihgeni@gmail.com

Abstrak - UNIKOM merupakan salah satu dari ribuan perusahaan swasta di Indonesia yang berbisnis di dalam bidang perguruan tinggi , setiap perguruan tinggi memiliki pencapaian bisnis salah satunya adalah menghasilkan sumber daya manusia yang memenuhi kebutuhan posisi kerja di perusahaan – perusahaan , di dalam mendapatkan pencapaian tersebut terdapat beberapa hal – hal yang mempengaruhi pencapaian tersebut , salah satunya adalah penilaian mutu akademik yang dilakukan oleh lembaga pemerintah BAN – PT yang disebut sebagai akreditasi , jika UNIKOM memiliki nilai akreditasi yang rendah , maka hal itu akan mengakibatkan kurangnya peminat masuknya mahasiswa baru , kurangnya penerimaan pekerjaan dari suatu perusahaan yang lulus dari UNIKOM , hal ini menurunkan pencapaian bisnis UNIKOM , dalam mengatasi hal ini salah satu dapat dilakukan suatu pencegahan terhadap mahasiswa yang masih studi , suatu pencegahan yang berasal dari suatu prediksi peluang kelulusan mahasiswa yang masih studi , dengan menggunakan teknik *data mining* dan menggunakan algoritma *naïve bayes classifier* , dari hasil pencegahan ini UNIKOM dapat mengambil langkah perubahan untuk mendapatkan pencapaian bisnis.

Kata Kunci : *Data Mining , Naïve Bayes Classifier , Prediksi Peluang Kelulusan Mahasiswa.*

Abstract - *Unikom is the one of the thousands of companies that do business in the environment of college in Indonesia , each college has a business goal , one of the goal which is to produce human resources that meet the needs of work positions in a companies , to obtaining these achievements , there are several things that affect to obtain these achievements , one of which is the assessment of academic quality carried out by the government agency BAN - PT which is referred to as accreditation , if UNIKOM has a low accreditation value , then it will result in a lack of interest in the entry of new students , lack of acceptance of work from a company that graduated from UNIKOM , this reduces UNIKOM's business goal , in overcoming this one prevention can be carried out by students who are still studying , a prevention that comes from a prediction of the chances of graduating students still in study , by using data mining techniques and using naïve bayes classifier algorithms , from the results of this prevention UNIKOM can make a steps of change to get business goal.*

Keywords : *Data Mining , Naïve Bayes Classifier , Student Graduation Prediction.*

I. PENDAHULUAN

Universitas Komputer Indonesia (UNIKOM) adalah salah satu Universitas (Swasta) dari ribuan Universitas Negeri maupun Swasta yang ada di Indonesia , UNIKOM merupakan perusahaan bisnis yang bergerak dalam bidang pendidikan perguruan tinggi , yang berikut serta dalam menghasilkan sumber daya manusia yang berkeahlian dan berpengalaman untuk siap mengisi posisi kerja perusahaan – perusahaan industri yang ada , terdapat banyak hal – hal yang dihadapi oleh UNIKOM dalam menjalankan proses bisnisnya , terdapat persaingan diantara Universitas lainnya , kualitas / nilai akreditasi Universitas , regulasi / peraturan pemerintah terhadap pendidikan tinggi di Indonesia , dan lain – lainnya , dari hal – hal tersebut berdampak terhadap proses bisnis UNIKOM yang mengakibatkan tujuan bisnis UNIKOM kurang tercapai.

Akreditasi merupakan suatu sertifikat yang diberikan oleh lembaga pemerintah BAN-PT kepada universitas yang merupakan nilai mutu pendidikan atau akademik dari suatu universitas, proses penilaian akreditasi memiliki beberapa faktor, salah satunya ialah melihat perkembangan statistik mahasiswa yang ada di suatu universitas, apakah mahasiswa dari suatu universitas memiliki IPK yang tinggi, apakah mahasiswa dari suatu universitas memiliki suatu sertifikat yang berkompeten, apakah suatu universitas memiliki data wisudawan yang bekerja di perusahaan yang besar, dan lain – lainnya.

UNIKOM telah ada sejak tahun 2000, UNIKOM telah banyak berhasil memberikan gelar Sarjana dan Magister kepada mahasiswa – mahasiswanya, namun dalam menjalankan proses bisnis UNIKOM, terdapat regulasi penilaian mutu universitas yang dibuat oleh BAN-PT yang disebut akreditasi, kualitas mutu akademik mahasiswa – mahasiswa UNIKOM termasuk dalam penilaian akreditasi, jika UNIKOM memiliki kualitas mutu akademik mahasiswa yang rendah hal itu mengakibatkan turunnya pencapaian tujuan bisnis yang ada di UNIKOM, salah satunya adalah berkurangnya mahasiswa yang mendaftarkan diri ke UNIKOM dan juga berkurangnya penerimaan lowongan kerja dari suatu perusahaan yang lulus dari UNIKOM.

Dalam data wisudawan jurusan sistem informasi UNIKOM pada 5 tahun terakhir diketahui bahwa terdapat 65% wisudawan lulus tepat waktu wisuda pada tahun 2018, terdapat 26% wisudawan lulus tepat waktu wisuda pada tahun 2017, terdapat 82% wisudawan lulus tepat waktu wisuda pada tahun 2016, terdapat 53% wisudawan lulus tepat waktu wisuda pada tahun 2015, dan terdapat 40% wisudawan lulus tepat waktu wisuda pada tahun 2014, dalam penilaian akreditasi jika suatu universitas memiliki total jumlah data wisudawan lulus tepat waktu $\geq 50\%$, maka diberi nilai 4 atau sangat baik, namun pada data wisudawan jurusan sistem informasi UNIKOM dari tahun ke tahun terdapat peningkatan dan penurunan pada data wisudawan, jika hal tersebut terus berlanjut tanpa ada suatu langkah perubahan, maka hal itu menambah kemungkinan untuk turun nilai mutu akademik sistem informasi UNIKOM, yang mengakibatkan turunnya tujuan bisnis UNIKOM, dalam mengatasi hal tersebut salah satunya bisa dilakukan suatu pencegahan pada mahasiswa yang masih studi, dengan cara mengetahui besar peluang kelulusan pada mahasiswa yang masih studi, yaitu menggunakan teknik *data mining* dan menggunakan algoritma *naïve bayes classifier* yang menghasilkan suatu presentase peluang kelulusan kepada mahasiswa yang masih studi dengan matakuliah sebagai parameter perhitungan presentase peluang kelulusan, dari hasil teknik ini UNIKOM dapat mengambil langkah perubahan terhadap matakuliah – matakuliah yang memiliki pengaruh besar terhadap kelulusan mahasiswa, dari hasil pencegahan ini UNIKOM dapat meningkatkan nilai mutu akademik jurusan sistem informasi UNIKOM, yang dapat meningkatkan pendapatan dari tujuan bisnis UNIKOM.

II. KAJIAN PUSTAKA

2.1. Universitas.

Sebuah lembaga / pendidikan tinggi yang menyediakan fasilitas untuk pengajaran dan penelitian dan berwenang untuk memberikan gelar akademik, khususnya satu terdiri dari divisi sarjana yang memberikan gelar sarjana dan divisi pascasarjana yang terdiri dari sekolah pascasarjana dan sekolah profesional yang masing-masing dapat memberikan gelar master dan doktor [1].

2.2. Akademik.

Akademik suatu yang digunakan untuk menggambarkan hal-hal yang berhubungan dengan aktivitas yang dilakukan di sekolah, perguruan tinggi, dan universitas, terutama aktivitas yang melibatkan belajar dan bernalar daripada keterampilan praktis atau teknis [2].

2.3. Akreditasi.

Sertifikasi resmi yaitu sekolah atau kursus telah memenuhi standar yang ditetapkan oleh regulator eksternal [3].

2.4. Data Mining.

“ Data Mining “ adalah proses yang digunakan oleh perusahaan untuk mengubah data mentah menjadi informasi yang bermanfaat. Dengan menggunakan perangkat lunak untuk mencari pola dalam kumpulan data yang besar, perusahaan dapat belajar lebih banyak tentang pelanggan mereka untuk mengembangkan strategi pemasaran yang lebih efektif, meningkatkan penjualan dan mengurangi biaya. Penambahan data tergantung pada pengumpulan data yang efektif, pergudangan dan pemrosesan komputer. “Data Mining” juga dikenal sebagai penemuan data dan penemuan pengetahuan [4].

2.5. Naïve Bayes.

Klasifikasi Naïve Bayes adalah algoritma yang menggunakan teorema Bayes untuk mengklasifikasikan objek. Klasifikasi Naïve Bayes menganggap independensi yang kuat, atau Naïve antara atribut poin data. Penggunaan populer dari pengklasifikasi Naïve Bayes termasuk filter spam, analisis teks dan diagnosis medis. Klasifikasi ini banyak digunakan untuk pembelajaran mesin karena mudah diterapkan [5].

2.6. NetBeans IDE.

Netbeans IDE adalah sebuah lingkungan pengembangan sistem bagi pemrogram untuk menulis , mengompilasi , mencari kesalahan dan menyebarkan program yang ditulis dalam bahasa pemrograman Java , namun dapat mendukung bahasa pemrograman lainnya [6].

2.7. Pemrograman Berorientasi Objek.

Everything is Object , adalah konsep dasar pemrograman berorientasi objek. Dimana , file (dalam bentuk *class*) dapat diwujudkan menjadi satu atau lebih objek. Dalam sebuah *project* atau program , terdapat *class* dan *interface* yang saling terintegrasi. Keterintegrasian *class* diwujudkan dalam hirarki *class* , dimana *abstract class* menduduki posisi tertinggi dalam hirarki tersebut. *Class abstract* adalah kelas yang menggambarkan sistem / program secara keseluruhan dengan maksud menyederhanakan sistem tersebut [7].

III. METODE PENELITIAN

3.1. Analisis Kasus.

Di dalam meningkatkan nilai mutu akademik jurusan sistem informasi UNIKOM strata 1 dibutuhkan analisis kepada faktor – faktor yang dapat meningkatkan atau menurunkan hal tersebut , di dalam riset penelitian ini salah satu faktornya adalah jika jumlah data wisudawan jurusan sistem informasi UNIKOM strata 1 memiliki jumlah data wisudawan lulus tepat waktu $\geq 50\%$, maka di dalam penilaian akreditasi data wisudawan akan diberi nilai 4 atau sangat baik , jika jumlah data wisudawan lulus tepat waktu jurusan sistem informasi UNIKOM strata 1 $< 50\%$, maka nilai akan berkurang , jika jumlah data wisudawan lulus tepat waktu jurusan sistem informasi UNIKOM strata 1 dari tahun ke tahun menurun , maka akan menyebabkan turunnya pencapaian bisnis UNIKOM.

3.2. Pengumpulan Data.

Data riset penelitian mahasiswa wisudawan dan non wisudawan UNIKOM jurusan Sistem Informasi didapatkan dengan memperoleh izin dari pihak Sekretariat Jurusan Sistem Informasi UNIKOM dan Ketua Prodi Sistem Informasi Unikom.

Data yang dipakai di dalam riset penelitian ini adalah data – data nilai mahasiswa jurusan sistem informasi UNIKOM strata 1 dari tahun 2011 sampai dengan tahun 2017 dan data wisudawan dari tahun 2011 sampai dengan tahun 2018 , pengumpulan data ini di dapatkan dengan memperoleh izin dari pihak ketua prodi sistem informasi UNIKOM untuk meminta data yang di inginkan , setelah mendapat izin dari ketua prodi data yang di inginkan dapat diminta kepada sekretariat jurusan sistem informasi UNIKOM.

3.3. Analisis Data.

Data – data nilai yang di dapatkan memiliki banyak atribut , di dalam riset penelitian ini atribut dari data nilai yang digunakan dalam riset penelitian ini hanya atribut Nim , Matakuliah , dan Index Nilai , dan atribut yang digunakan pada data wisudawan hanya atribut Nim , Tahun Lulus , dan Semester Genap atau Semester Ganjil kelulusan , hal ini dilakukan karena banyak dari data nilai memiliki data yang tidak lengkap pada beberapa atribut.

3.4. Langkah penerapan algoritma *naïve bayes classifier*.

Di dalam penerapan algoritma *naïve bayes classifier* digunakan 10 matakuliah terpilih dengan jumlah data wisudawan lulus tepat waktu tertinggi pada matakuliah tersebut , dengan *class* dari algoritma *naïve bayes classifier* adalah lulus tepat waktu dan tidak lulus tepat waktu , di dalam penerapan ini data – data nilai dari data wisudawan digunakan sebagai *data training* , dan data – data nilai dari data mahasiswa yang masih studi digunakan sebagai *data test* , di dalam hasil dari penerapan algoritma ini jika hasil dari kalkulasi peluang kelulusan kepada mahasiswa yang masih studi memiliki peluang 0% atau *zero frequency* maka data hasil dari perhitungan *naïve bayes classifier* tidak ditampilkan.

IV. HASIL DAN PEMBAHASAN

4.1. Hasil.

| Angkatan | Total Wisudawan | Tepat Waktu | Tidak Tepat Waktu |
|----------|-----------------|---------------|-------------------|
| 2004 | 1 | 0 [0%] | 1 [100%] |
| 2005 | 0 | 0 [0%] | 0 [0%] |
| 2006 | 8 | 0 [0%] | 8 [100%] |
| 2007 | 60 | 15 [25%] | 45 [75%] |
| 2008 | 123 | 38 [30.894%] | 85 [69.106%] |
| 2009 | 170 | 69 [40.588%] | 101 [59.412%] |
| 2010 | 155 | 83 [53.548%] | 72 [46.452%] |
| 2011 | 234 | 194 [82.906%] | 40 [17.094%] |
| 2012 | 203 | 53 [26.108%] | 150 [73.892%] |
| 2013 | 189 | 124 [65.608%] | 65 [34.392%] |
| 2014 | 6 | 5 [83.333%] | 1 [16.667%] |
| 2015 | 1 | 1 [100%] | 0 [0%] |
| 2016 | 1 | 1 [100%] | 0 [0%] |
| Total | 1151 | 583 [50.652%] | 568 [49.348%] |

| DaftarMK | TW_A | TW_B | TW_C | TW_D | TW_E | NTW_A | NTW_B | NTW_C | NTW_D | NTW_E |
|----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| IS31272 | 0.038 | 0.045 | 0.019 | 0.004 | 0.001 | 0.027 | 0.037 | 0.033 | 0.008 | 0 |
| IS36388 | 0.066 | 0.033 | 0.004 | 0.001 | 0.001 | 0.026 | 0.042 | 0.019 | 0.006 | 0 |
| IS32278 | 0.05 | 0.033 | 0.015 | 0.004 | 0.002 | 0.027 | 0.055 | 0.023 | 0.01 | 0 |
| IS34285 | 0.041 | 0.036 | 0.024 | 0.001 | 0 | 0.022 | 0.04 | 0.045 | 0.005 | 0 |
| IS34377L | 0.057 | 0.03 | 0.011 | 0.001 | 0 | 0.054 | 0.036 | 0.008 | 0.001 | 0 |
| IS34379L | 0.01 | 0.01 | 0.071 | 0.005 | 0.002 | 0.007 | 0.017 | 0.064 | 0.009 | 0 |
| IS37474 | 0.048 | 0.042 | 0.007 | 0 | 0.001 | 0.031 | 0.053 | 0.011 | 0.002 | 0 |
| IS32172 | 0.04 | 0.046 | 0.008 | 0.001 | 0.002 | 0.041 | 0.048 | 0.005 | 0 | 0 |
| IS32277 | 0.04 | 0.026 | 0.028 | 0.001 | 0.002 | 0.02 | 0.036 | 0.032 | 0.005 | 0 |
| IS33280 | 0.049 | 0.039 | 0.005 | 0.001 | 0.002 | 0.033 | 0.048 | 0.014 | 0 | 0 |

| No | Nim | Lulus | Peluang 1 | Peluang 2 |
|----|----------|-------------------|-----------|-----------|
| 1 | 10511007 | Tidak Tepat Waktu | 56.358% | 43.642% |
| 2 | 10511012 | Tidak Tepat Waktu | 60.597% | 39.403% |
| 3 | 10511025 | Tidak Tepat Waktu | 61.051% | 38.949% |
| 4 | 10511042 | Tidak Tepat Waktu | 82.816% | 17.184% |
| 5 | 10511047 | Tidak Tepat Waktu | 69.766% | 30.234% |
| 6 | 10511096 | Tepat Waktu | 89.354% | 10.646% |
| 7 | 10511101 | Tidak Tepat Waktu | 67.545% | 32.455% |
| 8 | 10511124 | Tepat Waktu | 63.704% | 36.296% |
| 9 | 10511129 | Tepat Waktu | 75.166% | 24.834% |
| 10 | 10511151 | Tidak Tepat Waktu | 57.479% | 42.521% |
| 11 | 10511212 | Tepat Waktu | 90.458% | 9.542% |
| 12 | 10511223 | Tepat Waktu | 96.541% | 3.459% |
| 13 | 10511283 | Tidak Tepat Waktu | 88.502% | 11.498% |
| 14 | 10511308 | Tepat Waktu | 75.573% | 24.427% |
| 15 | 10511360 | Tepat Waktu | 96.757% | 3.243% |
| 16 | 10511418 | Tepat Waktu | 95.132% | 4.868% |
| 17 | 10511468 | Tidak Tepat Waktu | 63.895% | 36.105% |
| 18 | 10512004 | Tidak Tepat Waktu | 96.549% | 3.451% |
| 19 | 10512009 | Tepat Waktu | 72.137% | 27.863% |
| 20 | 10512015 | Tidak Tepat Waktu | 70.293% | 29.707% |
| 21 | 10512027 | Tidak Tepat Waktu | 96.187% | 3.813% |

| No | Nim | Lulus | Peluang 1 | Peluang 2 |
|-------|-------------------|-------------------|-----------|-----------|
| 78 | 10514111 | Tidak Tepat Waktu | 96.651% | 3.349% |
| 79 | 10514112 | Tidak Tepat Waktu | 63.714% | 36.286% |
| 80 | 10514114 | Tepat Waktu | 77.185% | 22.815% |
| 81 | 10514116 | Tepat Waktu | 95.791% | 4.209% |
| 82 | 10514118 | Tepat Waktu | 69.282% | 30.718% |
| 83 | 10514127 | Tidak Tepat Waktu | 54.704% | 45.296% |
| 84 | 10514130 | Tepat Waktu | 93.67% | 6.33% |
| 85 | 10514133 | Tidak Tepat Waktu | 98.196% | 1.804% |
| 86 | 10514134 | Tidak Tepat Waktu | 89.122% | 10.878% |
| 87 | 10514135 | Tidak Tepat Waktu | 61.726% | 38.274% |
| 88 | 10514137 | Tidak Tepat Waktu | 75.109% | 24.891% |
| 89 | 10514145 | Tepat Waktu | 69.982% | 30.018% |
| 90 | 10514146 | Tepat Waktu | 59.028% | 40.972% |
| 91 | 10514148 | Tepat Waktu | 78.249% | 21.751% |
| 92 | 10514165 | Tepat Waktu | 66.163% | 33.837% |
| 93 | 10514203 | Tepat Waktu | 96.987% | 3.013% |
| 94 | 10514236 | Tepat Waktu | 93.421% | 6.579% |
| 95 | 10514261 | Tepat Waktu | 83.432% | 16.568% |
| 96 | 10514269 | Tepat Waktu | 78.604% | 21.396% |
| Total | Tepat Waktu | 40.0 [41.667%] | 82.198% | 17.802% |
| Total | Tidak Tepat Waktu | 56.0 [58.333%] | 75.471% | 24.529% |

4.2. Pembahasan.

Dapat diketahui bahwa jumlah data prediksi pada *data training* cukup rendah, hal ini dikarenakan kurang banyaknya *input* data nilai pada *software*, seperti *data set* diatas bahwa dari seluruh data wisudawan tidak terdapat wisudawan yang lulus tidak tepat waktu yang tidak mendapatkan nilai E pada 10 matakuliah terpilih, salah satu kemungkinannya adalah kurang banyaknya *input* data nilai, dikarenakan sampel data nilai untuk setiap angkatan

terbatas hanya 4 kelas , hal itu akan sangat mengganggu perhitungan algoritma *naïve bayes classifier* , karena jika salah satu pada *data training* mendapatkan nilai E pada salah satu 10 matakuliah terpilih , maka kalkulasi akan peluang kelulusan dikalikan angka 0 , hal itu menyebabkan peluang lulus menjadi 100 % dan peluang tidak lulus 0 % , hal ini tidak diinginkan.

V. KESIMPULAN DAN SARAN

Kesimpulan : dapat dilihat bahwa dari hasil prediksi terdapat 56 dari 96 (58 %) mahasiswa terprediksi lulus tidak tepat waktu , dengan total peluang tidak lulus tepat waktu sebesar 75 % , jika tanpa adanya perubahan pada matakuliah jurusan sistem informasi UNIKOM , maka hal ini akan berpeluang menurunkan nilai akreditasi jurusan sistem informasi UNIKOM.

Saran : *Software* ini masih terdapat banyak kekurangan , salah satunya belum ada pengelompokkan hasil *data training* berdasarkan 10 matakuliah terpilih , hal ini belum dilakukan karena total *data training* yang di prediksi masih sedikit di bawah 100 data , pengelompokkan itu sangat membantu jika total dari *data training* sangat banyak.

DAFTAR PUSTAKA

- [1] American Dictionary / Kamus Bahasa Negara Amerika Serikat, Definisi kata Universitas, Diakses pada tanggal 25 / 12 / 2018,
Website : <https://www.merriam-webster.com/dictionary/university>
- [2] American Dictionary / Kamus Bahasa Negara Amerika Serikat, Definisi kata Akademik, Diakses pada tanggal 25 / 12 / 2018,
Website : <https://www.collinsdictionary.com/dictionary/english/academic>
- [3] American Dictionary / Kamus Bahasa Negara Amerika Serikat, Definisi kata Akreditasi, Diakses pada tanggal 25 / 12 / 2018,
Website : <https://en.oxforddictionaries.com/definition/accreditation>
- [4] Definisi istilah “ Data Mining “ dalam persepsi Bisnis,
Diulas oleh : Will Kenton , Biodata : <https://www.investopedia.com/contributors/53661/>
Diakses pada tanggal 25 / 12 / 2018,
Website : <https://www.investopedia.com/terms/d/datamining.asp>
- [5] American Dictionary / Kamus Bahasa Negara Amerika Serikat, Definisi istilah Naïve Bayes, Diakses pada tanggal 25 / 12 / 2018,
Website : <https://www.techopedia.com/definition/32335/naive-bayes>
- [6] Fauzan Rauf & Mauluddin Syahrul , Pemrograman Java Berbasis GUI Menggunakan Database MySQL : Megatama.
- [7] Bella Hardiyana, S.Kom., M.Kom & Julian Chandra Wibawa, S.Kom., M.Kom , Belajar Pemrograman Berorientasi Objek Dengan Bahasa Java Buku-2 : Megatama.