

APLIKASI PREDIKSI KELULUSAN MAHASISWA DENGAN METODE LINEAR DISCRIMINANT ANALYSIS

Sulistio

PT Adicipta Inovasi Teknologi (AdIns) Jakarta
email : sulistio.workmail@gmail.com

Abstrak

Aplikasi Prediksi Status Kelulusan adalah sebuah aplikasi yang memprediksi status kelulusan siswa. Aplikasi Prediksi Status Kelulusan Berbasis Web ini dibuat dengan menggunakan metode yang disebut analisis diskriminan linier. Metode ini digunakan untuk mengklasifikasikan objek dari objek lain ke kelas masing-masing dengan menggunakan atributnya. Metode analisis diskriminan linier ini digunakan pada modul prediksi yang memisahkan sekelompok siswa ke kelas masing-masing (lulus tepat waktu atau terlambat) dan kemudian menunjukkan hasilnya kepada pengguna. Data yang digunakan untuk input adalah nilai subjek dari masing-masing siswa. Subjek yang digunakan dalam aplikasi ini dipilih oleh kepala jurusan dengan 22 subjek yang dipilih. Untuk tujuan evaluasi, metode cross validation digunakan, dengan hasil akurasi rata-rata sebesar 97%.

Kata kunci: Koefisien Korelasi, validasi K-Fold Cross, Analisis Diskriminan Linier

Abstract

Graduation Status Prediction Application is an application that predict graduation status of student. This Web Based Graduation Status Prediction Application is created using method called linear discriminant analysis . This method is used to classified an object from other object to their respective classes by using their attributes. This linear discriminant analysis method is used on predicting module that separate a group of student to their respective classes (passed on time or late) and then showing the result to the user. The data used for input is a subject scores from each student. The subject used in this application is selected by the head major with 22 subject selected currently. For evaluation purpose, k-fold cross validation method is used, with the result of average accuracy as big as 97%.

Key words : Correlation Coefficient, K-Fold Cross validation, Linear Discriminant Analysis

1. PENDAHULUAN

Pesatnya kemajuan teknologi di dunia informasi dalam berbagai bidang telah membawa dampak besar bagi umat manusia. Pada bidang pendidikan contohnya, teknologi telah membantu kemudahan proses belajar mengajar bagi pengajar atau pelajar dengan membuat proses kerja menjadi lebih efisien, contohnya pelajar bisa melihat nilai atau mengambil mata

kuliah dimana saja secara online atau pengajar bisa mengambil dan membawakan materi dengan mudah melalui perangkat keras laptop atau komputer.

Contoh aplikasi akademis yang akan membantu dalam bidang pendidikan adalah aplikasi untuk melakukan prediksi status kelulusan seorang mahasiswa di awal masa perkuliahan, sehingga mahasiswa bisa menyadari hal yang diperlukan untuk membantu dirinya di semester lanjut dan membantu pekerjaan pihak yang berwenang dalam menangani mahasiswa yang dianggap akan mengalami kesulitan untuk lulus tepat waktu.

Dari hasil Tanya jawab diketahui bahwa untuk sekarang, belum ada sistem yang mengurus tentang prediksi kelulusan mahasiswa. Di fakultas teknologi informasi universitas tarumanagara, khususnya Penilaian mahasiswa yang kira-kira perlu mendapat bimbingan dilakukan dengan cara melihat jumlah IPK dan SKS yang dapat diambil mahasiswa di semester empat. Jika IPK dinilai terlalu rendah (dua ke bawah) dan SKS yang dapat diambil 10 atau kurang, maka mahasiswa akan dipanggil untuk menanyakan penyebab dan kesulitannya berkuliah. Setelah itu mahasiswa akan diberi solusi atau sugesti yang dapat membantunya.

Untuk itu dibuatlah sebuah aplikasi prediksi kelulusan mahasiswa yang dibuat untuk Dosen Ketua Jurusan Universitas Tarumanagara yang akan berfungsi untuk memprediksi status kelulusan mahasiswa (lulus tepat waktu atau lulus tidak tepat waktu) berdasarkan nilai mahasiswa dalam mata kuliah yang dianggap sulit dan/atau penting di semester 1 sampai 4 sehingga dosen ketua jurusan bisa mengambil keputusan dengan lebih cepat dan akurat. Aplikasi prediksi kelulusan ini akan menggunakan metode *Linear Discriminant Analysis* yaitu sebuah metode yang menggunakan nilai dari beberapa variable bebas untuk menentukan kecenderungan masuknya suatu objek ke dalam satu kelompok (variable terikat) yang telah ditentukan yaitu lulus tepat waktu dan tidak tepat waktu. Nama mahasiswa kemudian ditampilkan dengan prediksi dan skor determinan yang didapat dari proses perhitungan fungsi diskriminan. Hal ini diharapkan bisa mempermudah proses prediksi kelulusan untuk mahasiswa yang mengalami masalah dalam mengikuti beberapa pembelajaran, sehingga pihak fakultas bisa membantu dengan memanggil mahasiswa bersangkutan untuk keperluan konseling atau pencarian solusi.

2. DASAR TEORI

2.1 *Statistical Classification*

Statistical classification atau sering juga disebut sebagai *classification* adalah sebuah cara atau metode untuk mengidentifikasi kemana sebuah objek data dikelompokkan. Sebagai contoh adalah klasifikasi *email* ke dalam kategori *spam* atau bukan *spam*, mahasiswa ke dalam kategori lulus tepat waktu, atau tidak tepat waktu, dan sebagainya. *Classification* juga bisa digunakan untuk membangun sebuah model prediktif untuk memprediksi objek baru [1]. yang masuk ke dalam set data atau objek. Dalam *statistical classification* biasanya, tiap observasi dari individual atau objek baru akan dianalisa ke dalam suatu set variabel yang bisa dijelaskan (*explanatory variable*) atau fitur. Nilai dari fitur ini bisa berupa kategori (seperti golongan darah A, B, AB, atau O), urutan/*ordinal* (seperti besar, kecil, atau sedang), atau berupa angka (bilangan bulat atau real). Ada juga *classifier* yang bekerja dengan melakukan perbandingan similaritas observasi lalu dan sekarang. Salah satu cara untuk melakukan *classification* adalah dengan menggunakan *classifier algorithm*.

2.1.1. *Classifier Algorithm*

Classifier algorithm atau disebut juga *Classification algorithm* dan *classifier* merupakan algoritma yang banyak dipakai untuk keperluan klasifikasi pada bidang ilmu *machine learning*

atau statistik. *Classification algorithm* biasanya digunakan sebagai metode perhitungan untuk melakukan prediksi data baru dengan karakteristik data lama [2]. *Classifier algorithm* juga biasanya merujuk kepada rumus atau fungsi matematika yang juga mempunyai kemampuan prediksi dengan mengklasifikasi data baru ke dalam kelas tertentu. Beberapa contoh *classifier algorithm* adalah *linear classifier*, *support vector machine*, *quadratic classifier*, *kernel estimation*, *boosting*, *decision tree*, *neural networks*, dan *learning vector quantization*. Semua algoritma klasifikasi mempunyai keuntungan dan kerugian masing – masing, tergantung dari macam - macam factor seperti tipe data set, jumlah kelas dan lain - lain.

2.1.2. Linear Classifier

Pada *statistical classification*, target atau objektif dari klasifikasi adalah menentukan atau mengidentifikasi kemana sebuah objek dimasukkan dalam beberapa grup atau kelas. Hal ini dilakukan dengan menggunakan kombinasi linear dari beberapa karakteristik atau fitur sebuah objek yang diolah menjadi *linear predictor function* untuk setiap kelas yang ada. Setelah itu objek baru akan ditentukan untuk masuk ke kelas tertentu berdasarkan hasil skor dengan cara memasukkan karakteristik objek baru ke dalam fungsi *predictor*. Nilai fungsi yang didapat dari proses klasifikasi ini akan menentukan kelas dari objek setelah diinterpretasi. Salah satu keunggulan dari *linear classifier* adalah kecepatannya dalam melakukan klasifikasi terutama untuk objek yang hanya mempunyai dua kelas. *Linear classifier* bisa dicari dengan menggunakan *generative model* dan *discriminative model*, yang membedakan model *generative* dan *discriminative* adalah prosedur training yang digunakan untuk mencari fungsi diskriminan dan bagaimana hasil skor dari fungsi diskriminan tersebut di interpretasikan. Contoh algoritma yang digunakan dalam *discriminative model* adalah metode *linear discriminant analysis*.

2.1.3. Correlation

Correlation atau sering juga disebut *correlation coefficient* (koefisien korelasi) adalah nilai yang digunakan untuk mengukur relasi antara dua pasang set data. Nilai dari korelasi ini bisa dikategorikan sebagai berikut :

- Korelasi positif – jika satu variabel naik, maka variabel lainnya juga akan ikut naik, dan begitupun sebaliknya;
- Korelasi negative – jika satu variabel naik, maka variabel lainnya turun, dan begitupun sebaliknya;
- Tidak ada Korelasi – kenaikan variabel tidak mempengaruhi variabel lainnya sama sekali.

Korelasi positif bisa di kelompokkan lagi menurut jumlah nilainya sebagai berikut:

- -1.00 : relasi *linear* negatif sempurna;
- -0.99 sampai -0.70 : relasi *linear* negatif kuat;
- -0.69 sampai -0.50 : relasi *linear* negatif sedang;
- -0.49 sampai -0.01 : relasi *linear* negatif lemah;
- 0 : tidak ada relasi *linear*.
- 0.01 sampai 0.49 : relasi *linear* positif lemah;
- 0.50 sampai 0.69 : relasi *linear* positif sedang;
- 0.70 sampai 0.99 : relasi *linear* positif kuat;
- 1.0 : relasi *linear* positif sempurna.

Langkah menghitung korelasi koefisien adalah sebagai berikut :

1. Hitung total baris kolom x atau kolom y (n)
2. Hitung total dari nilai kuadrat pada kolom x $(\sum x^2)$

3. Hitung total dari nilai kuadrat pada kolom y $(\sum y^2)$
4. Hitung kuadrat dari nilai total pada kolom x $(\sum x)^2$
5. Hitung kuadrat dari nilai total pada kolom y $(\sum y)^2$
6. Hitung total dari nilai x dikali y $(\sum xy)$
7. Hitung $S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n}$ (1)
8. Hitung $S_{yy} = \sum y^2 - \frac{(\sum y)^2}{n}$ (2)
9. Hitung $S_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n}$ (3)
10. Hitung koefisien korelasi dengan rumus $\hat{p} = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$ (4)

2.1.4. Linear Discriminant Analysis

Linear discriminant analysis adalah generalisasi dari Fisher linear discriminant yang diciptakan oleh Ronald A. Fisher pada tahun 1936 untuk melakukan berbagai pengukuran untuk masalah Taxonomi. *Linear discriminant analysis* sedikit berbeda dari fisher linear discriminant dimana *linear discriminant analysis* mengasumsikan variabel yang digunakan berdistribusi normal dan mempunyai kovarians dengan kelas yang sama. Tujuan dari *linear discriminant analysis* (LDA) adalah mengklasifikasikan objek ke dalam beberapa kelas berdasarkan fitur yang menggambarkan objek tersebut, di dalam *linear discriminant analysis* objek mempunyai dua variable yaitu variable kelas/terikat (dependent variable) dan variable atribut/bebas (independent variable), variable terikat mempunyai ikatan dengan variable bebas yang menggambarkan variable tersebut. Variable bebas nantinya akan digunakan untuk menentukan kombinasi linear dari objek tersebut. LDA bekerja dengan menggunakan analisa matriks penyebaran yang bertujuan menemukan proyeksi optimal sehingga dapat memproyeksikan data input pada ruang dengan dimensi yang lebih kecil dimana semua pola (pattern) dapat dipisahkan semaksimal mungkin [3]. Dalam LDA variable dependent merupakan kelas dari object yang biasanya mempunyai nilai nominal / nama dari kelas (lulus/tidak lulus) dan variable independent merupakan fitur yang menggambarkan object tersebut biasanya bernilai skalar. Sebelum melakukan prediksi, LDA membutuhkan fase training untuk menentukan fungsi diskriminan. Fase training ini membutuhkan objek yang telah terklasifikasi beserta sejumlah variabel fitur/independen variable. Langkah dalam menghitung fase training dalam LDA adalah sebagai berikut :

1. Kelompokkan training data ke dalam matriks sejumlah kelasnya dinotasikan dengan X_i
Keterangan :
i = jumlah kelas.
2. Hitung matriks rata-rata dari tiap kelas (μ_i)
3. Hitung rata-rata global dari keseluruhan data (μ)
4. Hitung *mean corrected data* (X_i^0) dengan mengurangkan setiap nilai pada X_i dengan nilai rata-rata global (μ)
5. Hitung covariance matriks (C_i) dari masing-masing X_i dengan rumus :

$$C_i = \frac{(x_i^0)^T x_i^0}{n_i}$$
 (5)
 Keterangan:
 n_i = banyak baris pada grup X_i
6. Hitung nilai covariance global matriks (C) dengan rumus :

$$C = \frac{1}{N} \sum_{i=1}^g n_i c_i$$
 (6)
 Keterangan:
 N = jumlah baris dari keseluruhan data.

7. Setelah itu hitung invers dari matriks C (C^{-1}).
8. Cari peluang dari tiap kelas (P_i).
9. Dengan begitu kita bisa menghitung fungsi diskriminan (f_i) Dengan rumus :
$$f_i = \mu_i C^{-1} X_k^T - \frac{1}{2} \mu_i C^{-1} \mu_i^T + \ln(p_i) \quad (7)$$
10. Setelah f_i setiap kelas diketahui, akurasi fungsi diskriminan bisa dilakukan dengan melakukan pemetaan ulang tiap grup dalam training data dengan fungsi diskriminan f_i
11. Objek dari data training akan masuk ke dalam kelas tergantung dari nilai maksimum f_i

Setelah rumus diskriminan di ketahui dan di uji akurasinya, maka fase training bisa dinyatakan selesai, dan fungsi diskriminan ini sudah bisa digunakan untuk mengklasifikasikan objek baru ke dalam kelas yang ditentukan.

Dengan menghitung X_k^T dari objek baru dan kemudian memasukkannya ke dalam fungsi diskriminan masing – masing kelas. Objek baru akan di kelompokkan ke dalam kelas yang diwakili oleh fungsi diskriminan yang mempunyai nilai terbesar.

2.1.5. Cross validation

Cross validation merupakan teknik yang digunakan untuk melakukan evaluasi sebuah model atau fungsi. Proses validasi menggunakan *cross validation* dinilai lebih baik daripada evaluasi biasa, karena evaluasi biasa tidak memberikan indikasi seberapa baik sebuah model yang di training bisa melakukan prediksi atau perhitungan sebuah data set baru yang belum pernah digunakan. Cara untuk mengatasi masalah tersebut adalah dengan hanya menggunakan sebagian data yang digunakan untuk training sebagai *sample set* dan sisanya sebagai *training set*. Hal ini bisa membantu dalam mengukur performa sebuah model terhadap data trainingnya sendiri. Hal tersebut adalah ide kasar dari bermacam – macam variasi *cross validation*.

2.1.6. K-Fold Cross validation

K-Fold Cross Validation merupakan salah satu variasi dari *cross validation*. Cara perhitungan *K-fold cross validation* sama dengan ide kasar dari *cross validation*, namun sebelum validasi dilakukan, data set yang digunakan akan dibagi terlebih dahulu kedalam k subset, yang kemudian dihitung persentase keberhasilan model yang dibuat dari subset $k-1$ dalam memprediksi salah satu subset k . Proses ini diulang sebanyak k kali untuk menentukan rata rata keberhasilan beserta standar deviasi dari keberhasilan perhitungan model. Keunggulan dari model ini dibandingkan dengan model *cross validation* lainnya adalah, tidak peduli bagaimana data dibagi, setiap subset k akan divalidasi sebanyak satu kali dan juga akan dijadikan data training sebanyak $k-1$ kali. Sehingga hasil evaluasi menjadi lebih baik, meskipun perhitungan menjadi lebih lambat karena pengulangan sebanyak k .

3. PEMBAHASAN

Pengujian data dilakukan ketika pengujian modul selesai dilakukan. Pengujian data ini dilakukan dengan menggunakan data training dari pihak Fakultas Teknologi Informasi Universitas tarumanagara dengan jumlah data sebanyak 462 mahasiswa dari angkatan 2008-2012 yang telah di ketahui status kelulusannya dengan cara melihat pengambilan skripsi pada semester 8 atau 7, jika mahasiswa mengambil skripsi lewat dari semester tersebut, maka mahasiswa itu dinyatakan tidak lulus tepat waktu, dan begitupun sebaliknya jika mahasiswa

mengambil skripsi pada semester 8 atau 7, maka mahasiswa dianggap lulus tepat waktu. Informasi data sampel dan data training untuk pengujian dilakukan menggunakan data yang didapat dari pihak Universitas Tarumanagara. Pengujian data ini dilakukan dengan menggunakan metode *k-fold cross validation*, yaitu proses validasi data dengan melakukan prediksi terhadap subset dari dataset yang telah dibagi sebanyak *k*.

Cara penentuan skor prediksi didapat dengan menggunakan metode analisa diskriminan. Sebelum menentukan metode diskriminan, akan dilakukan terlebih dahulu perhitungan korelasi antara masing masing mata kuliah dengan status kelulusan, hal ini dilakukan karena ada beberapa mata kuliah yang sebenarnya bisa diwakilkan oleh mata kuliah lainnya tanpa mengurangi akurasi fungsi diskriminan. Mata kuliah yang digunakan antara lain adalah Kalkulus 1, Algoritma 1, Manajemen Organisasi, Pengenalan Komputer, Matematika Diskrit, Statistika 1, Sistem Informasi, Algoritma 2, Sistem Digital, Sistem Operasi, Interaksi Manusia dan Komputer, Aljabar Linear, Fisika Mekanika, Analisis Algoritma, Statistika 2, Pengantar Intelegensia Buatan, Teori Graf, Teknik Simulasi, Kalkulus 2, Struktur Data, Jaringan Komputer, dan Pemrograman GUI dengan NET.

Berikut adalah tabel korelasi mata kuliah dengan status kelulusan Mahasiswa :

Tabel 1. Korelasi tiap mata kuliah dengan status kelulusan

Mata Kuliah	Skor Korelasi
KALKULUS I	0.5445032135827
ALGORIT.& PEMROGR.C++ DSR	0.50358538061008
MANAJEMEN & ORGANISASI	0.44074879549712
PENGENALAN KOMPUTER	0.46778183193119
MATEMATIKA DISKRIT	0.58069515420779
STATISTIKA I	0.52758493961934
SISTEM INFORMASI	0.39699084456906
ALGORIT.& PEMROGR.C++LANJ	0.46461606323931
SISTEM DIGITAL	0.62480328040473
SIS.OPERASI & PRAK.LINUX	0.5300054892236
INTER. MANUSIA & KOMPUTER	0.46145780058079
ALJABAR LINIER	0.54769188568976
FISIKA MEKANIKA	0.54207782307272
ANALISIS ALGORITMA	0.5781577210263
STATISTIK II	0.58157295324348
PENG. INTELEGENSI BUATAN	0.55029620333068
TEORI GRAF	0.51350574091775
TEKNIK SIMULASI	0.48817652769498

STRUKTUR DATA	0.578078284129
JAR.KOMP.I & PRAK. LAN	0.60236288538581
PEMROGRAMAN GUI DGN NET	0.43156533069066
KALKULUS II	0.567128169721

Berdasarkan perhitungan korelasi, mata kuliah yang akan dipilih adalah mata kuliah dengan nilai korelasi ≥ 0.5 atau korelasi positif sedang yaitu Kalkulus 1, Algoritma 1, Matematika Diskrit, Statistika 1, Sistem Digital, Sistem Operasi, Aljabar Linear, Fisika Mekanika, Analisis Algoritma, Statistika 2, Pengantar Intelegensia Buatan, Teori Graf, Kalkulus 2, Struktur Data, dan Jaringan Komputer.

Setelah melakukan perhitungan korelasi, *k-fold cross validation* untuk fungsi diskriminan bisa dihitung, Hasil validasi *k-fold cross validation* fungsi diskriminan dengan nilai tengah mean yaitu data akurasi setiap iterasi beserta mean akurasi keberhasilan dan standar deviasinya bisa dilihat pada Tabel berikut.

Tabel 1. Persentase kebenaran dari *k-fold cross validation* dengan $K = 10$ dan nilai tengah mean

Iterasi	persentase kebenaran
1	0.98076923076923
2	0.97596153846154
3	0.98076923076923
4	0.97355769230769
5	0.97596153846154
6	0.97115384615385
7	0.96394230769231
8	0.98557692307692
9	0.96634615384615
10	0.97836538461538

Dari hasil evaluasi dengan menggunakan *k-fold cross validation* di atas, bisa di ketahui bahwa model diskriminan yang dibuat dengan data angkatan 2008–2012 yang status kelulusannya telah di ketahui berdasarkan waktu pengambilan skripsi (semester 8 ke bawah tepat waktu, jika lewat dari 8 semester berarti tidak tepat waktu) mampu memprediksi status kelulusan mahasiswa dengan rata-rata tingkat akurasi yang tinggi yaitu 0.97524038461538 dan dengan standar deviasi yang rendah yaitu sebesar 0.0009882117688026 atau bisa diartikan variasi dari tiap akurasi tidak berbeda jauh untuk masing masing pengujian. Sehingga model ini diperkirakan bisa digunakan untuk memprediksi status kelulusan mahasiswa dengan cukup tepat.

Selain menggunakan *k-fold cross validation*, dilakukan juga evaluasi terhadap kondisi sebenarnya. Sebelum melakukan evaluasi, data terlebih dahulu diambil secara acak dari data mahasiswa angkatan 2008-2012, akan tetapi diambil lebih dominan angkatan yang telah lulus semua yaitu angkatan 2008-2011. Setelah itu barulah dilakukan prediksi dengan fungsi diskriminan yang telah dibuat.

Berkut adalah tabel yang menampilkan hasil test dengan data yang sebenarnya

Tabel 3. Tabel evaluasi hasil prediksi kelulusan terhadap data yang sebenarnya

NIM	F1(TP)	F2(TTP)	Prediksi	Status Asli
53508xx	5.760731	4.771442	TP	TP
53508xx	3.149358	2.568873	TP	TP
53508xx	3.24939	2.439733	TP	TP
53508xx	4.46365	3.835099	TP	TP
53508xx	3.447809	2.515659	TP	TP
53508xx	4.124697	3.993859	TP	TP
53508xx	3.008789	3.757178	TTP	TTP
53508xx	2.97147	2.773845	TP	TP
NIM	F1(TP)	F2(TTP)	Prediksi	Status Asli
53508xx	-4.91711	-1.1902	TTP	TTP
53508xx	2.507236	3.512495	TTP	TTP
53508xx	3.376343	3.107528	TP	TTP
53508xx	3.340349	2.322289	TP	TP
53509xx	-4.51249	-0.79623	TTP	TTP
53509xx	8.309802	10.69586	TTP	TTP
53509xx	-4.36566	-0.70433	TTP	TTP
53509xx	4.545343	7.230996	TTP	TTP
53509xx	3.033441	3.668657	TTP	TTP
53509xx	4.069885	4.155745	TTP	TTP
53509xx	4.05622	5.0092	TTP	TTP
53509xx	1.957056	1.708298	TP	TP
53509xx	5.458577	6.435361	TTP	TTP
53509xx	-3.35261	0.249826	TTP	TTP
53509xx	4.500802	3.599059	TP	TP
53509xx	-0.95687	1.602285	TTP	TTP
53509xx	4.566377	7.602546	TTP	TTP
53509xx	6.988382	7.003502	TTP	TP
53509xx	0.472376	-0.33456	TP	TTP
53509xx	2.431116	3.488023	TTP	TTP
53509xx	-5.20836	-1.90614	TTP	TTP
53509xx	5.675335	5.626672	TP	TP
53509xx	1.550949	2.219866	TTP	TTP
53509xx	5.851963	4.31312	TP	TP
53509xx	-0.39618	2.462287	TTP	TTP
53510xx	6.805087	5.505548	TP	TP
53510xx	4.804802	5.648488	TTP	TTP
53510xx	1.709683	2.935271	TTP	TTP
53510xx	6.802169	6.197417	TP	TP
53510xx	3.500307	3.683374	TTP	TP
53510xx	-2.9626	0.608214	TTP	TTP
53510xx	6.561163	5.253956	TP	TP
53510xx	6.851601	6.050481	TP	TP
53510xx	3.244781	6.087617	TTP	TTP
53510xx	2.931207	2.502388	TP	TP
53510xx	3.505926	3.576274	TTP	TTP
53510xx	-2.23736	1.728325	TTP	TTP
53510xx	3.895822	3.446088	TP	TTP

53510xx	4.275853	3.472616	TP	TP
53510xx	4.70442	3.946283	TP	TP
53510xx	6.802169	6.197417	TP	TP
53510xx	3.500307	3.683374	TTP	TP
53510xx	-2.9626	0.608214	TTP	TTP
53510xx	6.561163	5.253956	TP	TP
53510xx	6.851601	6.050481	TP	TP
53510xx	3.244781	6.087617	TTP	TTP
53511xx	1.116941	1.11778	TTP	TTP
53511xx	5.629925	7.209402	TTP	TTP
53508xx	5.760731	4.771442	TP	TP
53508xx	3.149358	2.568873	TP	TP
53508xx	3.24939	2.439733	TP	TP
NIM	F1(TP)	F2(TTP)	Prediksi	Status Asli
53508xx	4.46365	3.835099	TP	TP
53508xx	3.447809	2.515659	TP	TP
53508xx	4.124697	3.993859	TP	TP
53508xx	3.008789	3.757178	TTP	TTP
53508xx	2.97147	2.773845	TP	TP
53508xx	4.966252	4.346115	TP	TP
53508xx	-4.91711	-1.1902	TTP	TTP
53511xx	-3.85737	-0.2147	TTP	TTP
53511xx	3.551339	3.257084	TP	TP
53511xx	2.888152	4.052451	TTP	TTP
53511xx	0.903208	0.839834	TP	TP
53511xx	4.059372	4.109303	TTP	TP
53511xx	4.059372	4.109303	TTP	TP
53511xx	3.511761	3.351048	TP	TP
53511xx	4.227187	4.105138	TP	TTP
53511xx	2.589827	3.168644	TTP	TTP
53511xx	4.749142	3.709915	TP	TP
53511xx	5.891646	6.347312	TTP	TP
53511xx	3.048957	3.248125	TTP	TTP
53511xx	6.798641	5.445576	TP	TP
53511xx	6.648056	5.322937	TP	TP
53511xx	6.985704	5.657514	TP	TP
53511xx	5.925863	5.03334	TP	TP
53512xx	6.226661	6.145351	TP	TP
53512xx	5.153755	4.195419	TP	TTP
53512xx	4.776775	3.603578	TP	TP
53512xx	7.482203	7.017821	TP	TP
53512xx	7.031486	5.932544	TP	TP
53512xx	3.918742	4.806909	TTP	TTP
53512xx	6.15393	5.063247	TP	TP
53512xx	2.266704	3.190072	TTP	TP
53512xx	4.671987	4.154253	TP	TP
53512xx	6.688716	5.526372	TP	TP
53512xx	5.45342	5.834071	TTP	TTP
53512xx	5.934676	4.509174	TP	TP
53512xx	7.334315	6.64871	TP	TP
53512xx	6.694664	6.062424	TP	TP
53512xx	5.595951	4.277811	TP	TP

53512xx	3.343784	4.068016	TTP	TTP
53512xx	5.538413	4.270797	TP	TP

Keterangan: *TP : Lulus tepat waktu, TTP : Lulus tidak tepat waktu

Jumlah individu yang digunakan untuk proses *testing* sebanyak 100 dengan kegagalan prediksi sebanyak 7 individu (baris dengan *shading* gelap) sehingga kemungkinan terjadi kegagalan sebesar 7/100 atau 0.07 dengan demikian persentase keberhasilan adalah 93%.

4. KESIMPULAN

1. Aplikasi ini dapat melakukan prediksi status kelulusan mahasiswa yang cukup akurat yaitu lebih dari 90% untuk evaluasi terhadap kondisi sebenarnya, maupun rata rata dari tiap iterasi *k-fold cross validation*.
2. Aplikasi ini telah dapat berjalan dengan baik sesuai dengan perancangan dan kebutuhan prediksi status mahasiswa.

DAFTAR PUSTAKA

- [1] Minsoo Kim, "Statistical Classification", Pomona College Paper, 2010 pg. 5
- [2] Patrick Ozer, "Data Mining Algorithms for Classification", BSc Thesis Artificial Intelligence, 2008 pg. 2
- [3] Ghazali, Eko K. Subha, Galuh M., M. Burhannudin, M.Izzun Niam, "Aplikasi Kematangan Tomat Berdasarkan Warna dengan Metode Linear Discriminant Analysis (LDA)", Universitas Brawijaya, Malang, 2013 Pg. 2