

Prediksi Kelangsungan dan Keberhasilan Studi Mahasiswa di Politeknik Negeri Batam

Hilda Widyastuti
Jurusan Teknik Informatika, Politeknik Negeri Batam
Batam, Indonesia
email: hilda@polibatam.ac.id

Abstrak—Politeknik Negeri Batam telah menetapkan aturan untuk penghentian studi secara dini bagi mahasiswa yang dianggap tidak berpotensi menyelesaikannya berdasarkan hasil studi di tahun pertama (IPK Tingkat 1), yang tertuang di peraturan akademik pasal 22 ayat 4. Penelitian ini bertujuan untuk mendukung keberadaan aturan tersebut menggunakan pendekatan klasifikasi dengan metode pohon keputusan (J48) dan pendekatan prediksi dengan metode regresi. Hasil penelitian menyatakan bahwa IPK tingkat 1 bisa digunakan untuk memprediksi kelangsungan dan keberhasilan kuliah mahasiswa, dan memprediksi IPK saat lulus. Hasil lainnya menyatakan pasal 22 ayat 4 harus disempurnakan supaya bisa digunakan secara efektif sebagai dasar penghentian dini mahasiswa yang tidak berpotensi dalam menyelesaikan pendidikannya.

Kata kunci—klasifikasi, pohon keputusan, prediksi, regresi, peraturan akademik

I. PENDAHULUAN

Politeknik Negeri Batam terletak di pulau Batam yang termasuk ke dalam kawasan perdagangan dan pelabuhan bebas dan juga merupakan kawasan terdepan dan terluar yang berbatasan langsung dengan perairan internasional. Dalam peraturan akademik Politeknik Negeri Batam [1] pasal 22 tentang penilaian prestasi mahasiswa, ayat 4, tertulis peraturan, “Khusus untuk mahasiswa tahun pertama, Indeks Prestasi Kumulatif (IPK) harus lebih besar atau sama dengan 1.5. Jika IPK kurang dari 1.5, mahasiswa tidak diperkenankan melanjutkan studi lagi di Politeknik Negeri Batam”. Pasal tersebut bertujuan untuk menghentikan secara dini mahasiswa yang tidak berpotensi menyelesaikan pendidikannya, berdasarkan hasil proses belajar-mengajar di tahun pertama (IPK tingkat 1). Pasal 22 ayat 4 tersebut mulai diberlakukan pada tahun ajaran 2012-2013.

Ada asumsi bahwa jika di tingkat 1 nilai mahasiswa rendah, maka nilai semester selanjutnya juga akan rendah. Kelompok mahasiswa tersebut biasanya tidak melanjutkan studinya atau jika tetap berusaha melanjutkan studinya, mereka mengalami kesulitan lulus kuliah atau tidak bisa memenuhi batas waktu studi yang telah ditetapkan, sehingga sebaiknya dihentikan studinya oleh kampus.

Penelitian ini bertujuan mendukung pasal 22 ayat 4 tersebut menggunakan pendekatan data mining. Rumusan masalahnya adalah :

1. Bagaimana memprediksi kelangsungan kuliah mahasiswa berdasarkan IPK tingkat 1?
2. Bagaimana memprediksi keberhasilan kuliah mahasiswa berdasarkan IPK tingkat 1?
3. Bagaimana memprediksi IPK saat lulus berdasarkan IPK tingkat 1?
4. Apakah pasal 22 ayat 4 peraturan akademik Politeknik Negeri Batam sudah efektif?
5. Berapa batas nilai yang tepat untuk menghentikan mahasiswa berdasarkan IPKnya?

II. PENELITIAN TERKAIT

Penggunaan data mining dengan metode klasifikasi dan prediksi untuk memprediksi keberhasilan studi mahasiswa sudah dilakukan di beberapa penelitian sebelumnya. Jika dilihat dari ruang lingkup tujuan penelitian, terdapat tiga tingkatan, yaitu : (1) penelitian dengan ruang lingkup keberhasilan studi pada perguruan tinggi atau sekolah, (2) penelitian dengan ruang lingkup keberhasilan menyelesaikan mata kuliah, (3) penelitian dengan ruang lingkup keberhasilan menyelesaikan sebuah tes pada suatu mata kuliah.

Contoh penelitian dengan ruang lingkup keberhasilan studi pada perguruan tinggi atau sekolah adalah penelitian untuk memprediksi *drop-out* berdasarkan kebiasaan sosialisasi mahasiswa [2], memprediksi keberhasilan mahasiswa yang menempuh pendidikan jarak jauh dengan e-learning [3], memprediksi salah orientasi pada lingkungan belajar online [4], memprediksi kegagalan sekolah [5], memprediksi performansi mahasiswa level master berdasarkan hasil kuliah saat sarjana [6], memprediksi *drop out* mahasiswa. Model yang ditemukan pada contoh terakhir ini digunakan untuk mendukung saran akademik, apakah seorang mahasiswa sebaiknya melanjutkan studinya atau tidak, berdasarkan nilai, informasi dosen pengajar semester 1, dan mentor mahasiswa [7].

Contoh penelitian dengan ruang lingkup keberhasilan menyelesaikan mata kuliah adalah penelitian untuk

memprediksi kesuksesan kuliah matematika berdasarkan perilaku pemrograman mahasiswa [8] dan memprediksi nilai akhir berdasarkan tingkat partisipasi mahasiswa pada forum diskusi di internet [9]. Sedangkan contoh penelitian dengan ruang lingkup keberhasilan menyelesaikan sebuah tes pada suatu mata kuliah adalah penelitian untuk memprediksi hasil tes berdasarkan interaksi antara mahasiswa dengan aplikasi *Cognitive Tutor* dan menggunakan data *Intelligent tutoring system (ITS)* untuk memprediksi hasil ujian [11]. Penelitian yang kami lakukan berada di ruang lingkup yang pertama, yaitu penelitian dengan ruang lingkup memprediksi keberhasilan studi mahasiswa pada perguruan tinggi atau sekolah.

Penelitian-penelitian yang terkait menggunakan beragam metode klasifikasi dan prediksi, antara lain *Bayesian network* [8] [9] [3], *RandomForest Regression* [4] [6] [7] [9], *JRip* [5] [7] [9], *NNge* [5] [9] [3], *RandomTree* [5], *J48* [11] [5] [7] [3] [9], *Logistic regression model* [7] [9] [10], *Prism* [5], *CART* [7], *OneR* [5] [3], *DTNB* [9], *Ridor* [9] [5], *ADTree* [9], [5] [9], *MultilayerPerceptron* [9], *SMO* [9], *RTree* [3]. Penelitian yang kami lakukan menggunakan metode J48 dan regresi linear. Penelitian yang kami lakukan dan beberapa penelitian terkait [5] [14] [7] menggunakan perangkat lunak yang sama, yaitu software Weka.

III. LANDASAN TEORI

A. Peraturan Akademik

Peraturan akademik Politeknik Negeri Batam [1] Pasal 21 tentang Yudisium dan Predikat Kelulusan ayat 3, 4, 5, dan 6 yang berbunyi :

- Predikat kelulusan yang diberikan kepada lulusan program pendidikan di Politeknik Negeri Batam dapat berupa *dengan pujian, sangat memuaskan, dan memuaskan.*
- Predikat kelulusan *dengan pujian* diberikan jika $IP > 3,50$ dan menyelesaikan pendidikannya sesuai dengan waktu tempuh normal. Jika menyelesaikan pendidikan lebih dari waktu tempuh normal mendapat predikat kelulusan *sangat memuaskan.*
- Predikat kelulusan *sangat memuaskan* diberikan jika $2,75 < IP \leq 3,50$.
- Predikat kelulusan *memuaskan* diberikan jika $2,0 \leq IP \leq 2,75$.

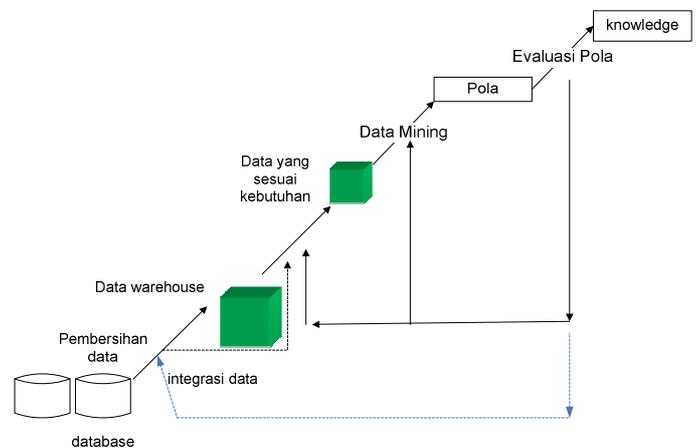
B. Data Mining

Data mining merupakan salah satu langkah dalam menemukan pengetahuan (*knowledge discovery*). Berdasarkan [12], langkah-langkah dalam menemukan pengetahuan dapat dilihat di gambar 1, meliputi :

- Pembersihan data, yaitu proses membuang *noise* dan data yang tidak konsisten. Data *noise* adalah data yang mengandung *error* atau data yang menyimpang dari yang diharapkan. Data *noise* terjadi karena kesalahan komputer atau manusia. Contoh : untuk menunjukkan jenis kelamin, digunakan L dan P. Ternyata terdapat data I. Sedangkan data yang tidak konsisten adalah data yang mengandung ketidakcocokan. Contohnya, ketidakcocokan kode

departemen yang digunakan untuk mengkategorisasikan suatu item. Di record pertama, Bill bekerja di departemen SDM, sedangkan di record kedua William bekerja di departemen HRD.

- Integrasi data, yaitu menggabungkan data dari berbagai sumber data.
- Seleksi data, mengambil data yang relevan dengan kepentingan analisis saja.
- Transformasi data, yaitu mengubah atau menggabungkan data ke bentuk-bentuk yang cocok untuk keperluan data mining, misalnya dengan operasi *summary*.
- Data mining, merupakan metode berintelegensia yang diterapkan dalam rangka mengekstrak pola-pola data.
- Evaluasi pola, yaitu mengidentifikasi pola yang menarik, yang merepresentasikan pengetahuan berdasarkan ukuran-ukuran tertentu.
- Menampilkan pengetahuan, yaitu menggunakan visualisasi dan teknik representasi pengetahuan untuk menampilkan pengetahuan hasil proses mining kepada pengguna.



Gambar 1. Langkah-langkah menemukan pengetahuan pada data mining

C. Klasifikasi dan Prediksi

Klasifikasi adalah penentuan suatu data ke kelas atau kelompok atau klasifikasi tertentu. Misalnya ada seorang siswa SMA yang berumur 17 tahun. Siswa SMA itu dikelompokkan sebagai remaja. Sedangkan definisi lengkap klasifikasi meliputi pembuatan model berdasarkan himpunan pelatihan dan menggunakan model tersebut untuk mengklasifikasikan data baru [12].

Prediksi digunakan untuk memperkirakan nilai yang kontinu. Misalnya prediksi gaji alumni Politeknik dengan pengalaman kerja 3 tahun. Metode yang digunakan untuk prediksi antara lain regresi, *back propagation*, *support vector machine*, *k-nearest-neighbor*, dan lain-lain. Pada penelitian ini metode yang dibahas adalah metode regresi.

Metode regresi digunakan untuk memodelkan hubungan antara satu atau lebih *independent variable* (*predictor variable*) dengan sebuah *dependent variable* (*response variable*). *Predictor variable* sudah diketahui, sedangkan yang ingin dicari nilainya adalah *response variable* [12]. Jenis-jenis

regresi adalah regresi linier, regresi linier berganda, dan regresi non linier misalnya regresi Poisson, regresi log-linear, dan regresi Polynomial. Bentuk umum fungsi regresi linier adalah :

$$Y = w_0 + w_1 x$$

Keterangan :

- w_0 dan w_1 : koefisien regresi
- x : masukan fungsi
- y : keluaran fungsi

D. K-fold Cross Validation

K-fold cross-validation adalah salah satu teknik untuk mengevaluasi keakuratan model, dengan ciri-ciri [12]:

1. Mempartisi data secara random ke dalam k buah himpunan/fold yaitu $D_1, D_2, ..D_k$. Setiap kelompok mempunyai jumlah yang hampir sama.
2. Pada perulangan i, gunakan D_i sebagai data uji dan himpunan lainnya sebagai data pelatihan
Contoh :
 - o Pada perulangan ke-1 : D_1 sebagai data uji dan D_2 s.d. D_k sebagai data pelatihan
 - o Pada perulangan ke-2 : D_2 sebagai data uji dan D_1, D_3 s.d. D_k sebagai data pelatihan
 - o dan seterusnya
3. Melakukan training dan pengujian sebanyak k kali
4. Menghitung keakuratan dengan rumus (1).

$$\text{keakuratan} = \frac{\text{jumlah hasil klasifikasi benar dari k iterasi}}{\text{total jumlah tuple}} \quad (1)$$

IV. PENYELESAIAN MASALAH

Eksperimen ke-1 menyelesaikan rumusan masalah ke-1, ke-2, ke-4, dan ke-5, sedangkan eksperimen ke-2 menyelesaikan rumusan masalah ke-3.

A. Eksperimen ke-1

Langkah-langkah eksperimen ke-1 adalah :

1. Pengumpulan data dan integrasi data

TABEL I. ATRIBUT TABEL

Nama	Deskripsi	Tipe data
NIM	Nomor induk mahasiswa	Text
Nama	Nama mahasiswa	Text
IPKTingkat1	IPK setelah menempuh semester 1 dan semester 2	Numerik
IPK	IPK setelah lulus	Numerik
LamaStudi	Lama mahasiswa belajar, dengan satuan semester	Numerik
Prodi	Program Studi (IF, EL, AK)	Text
Kelas	Program kuliah pagi atau kuliah malam	Text
Status	Lulus atau berhenti	Text
Grade	Posisi mahasiswa. Ada empat posisi yaitu gradesatu, gradedua, gradetiga, gradempat	Text

Data masukan untuk proses data mining adalah data nilai mahasiswa Politeknik Negeri Batam angkatan 2001 s.d. lulusan terakhir tahun 2010. Pada periode tersebut belum diterapkan pasal 22 ayat 4. Data diambil dari Sistem Informasi Politeknik Negeri Batam (SIMPOL). Data yang dibutuhkan disalin ke file MS Excel dengan atribut di tabel I. Jumlah data yang digunakan sebanyak 1.866 buah. Penelitian ini menggunakan software data mining WEKA versi 3.6.7. Supaya data file Excel bisa dibaca oleh WEKA, maka data yang berformat Excel ini akan dikonversi ke format CSV(*comma delimited*) [13].

2. Pembersihan data
Pembersihan data dilakukan dengan cara menghilangkan tanda titik(.), tanda petik('), tanda petik dua(""), tanda hubung(-), karena tanda-tanda tersebut menyebabkan file csv tidak bisa dibaca oleh WEKA.
3. Transformasi data
Pada saat penyalinan data ke format MS Excel, isi atribut IPKTingkat1, IPK yang bertipe numerik berubah menjadi teks, sehingga harus dilakukan perubahan ke format numerik lagi supaya bisa diolah lebih lanjut. Atribut IPK, lamaStudi, dan status adalah output yang dicapai oleh mahasiswa setelah dia menyelesaikan kuliahnya. Ketiga atribut tersebut berjenis output, sehingga dapat digabung menjadi satu.

Aturan penggabungannya berdasarkan Peraturan akademik Politeknik Negeri Batam 2012 Pasal 21 ayat 3, 4, 5, dan 6 [1] yaitu :

- IF IPK > 3.5 AND LamaStudi = 6 AND Status = "lulus" THEN Grade = "gradesatu"
- IF IPK > 3.5 AND LamaStudi > 6 AND Status = "lulus" THEN Grade="gradedua"
- IF IPK > 2.75 AND IPK <= 3.5 AND LamaStudi >= 6 AND Status = "lulus" THEN Grade="gradedua"
- IF IPK > 2.00 AND IPK <= 2.75 AND LamaStudi >= 6 AND Status = "lulus" THEN Grade = "gradetiga"
- IF Status = "berhenti" THEN Grade="gradempat"

4. Seleksi data

Tidak semua data digunakan pada data mining. Data dengan atribut IPK, LamaStudi, dan Status sudah ditransformasi ke atribut GradeLulus sehingga atribut IPK, LamaStudi, dan Status tidak diikutsertakan dalam proses data mining. Sedangkan atribut NIM, Nama, kelas, dan prodi juga tidak diikutsertakan dalam proses data mining, dengan alasan penelitian ini menginginkan hasil untuk ruang lingkup Politeknik Negeri Batam. Atribut tabel setelah menyelesaikan seleksi data ada di tabel II.

TABEL II. ATRIBUT TABEL

Nama	Deskripsi	Tipe data
IPKTingkat1	IPK setelah menempuh semester 1 dan semester 2	Numerik
Grade	Posisi mahasiswa	Text

5. Data mining
 Penelitian ini menggunakan metode klasifikasi J48 [13].
 Aturan yang terbentuk adalah :

```

IF IPKTingkat1 <= 2.76 THEN
  IF IPKTingkat1 <=1.9 THEN
    grade=" gradempat"
  ELSE
    IF IPKTingkat1 <=2.59 THEN
      grade=" gradetiga"
    ELSE
      grade=" gradedua"
    ELSE
      IF IPKTingkat1 <= 3.42 THEN
        grade=" gradedua"
      ELSE
        grade=" gradesatu"
  
```

6. Evaluasi pola
 Untuk mengevaluasi keakuratan model klasifikasi digunakan teknik *10-fold cross-validation*. Aturan yang terbentuk di langkah ke-5 mempunyai tingkat keakuratan 73.91%.
7. Menampilkan pengetahuan
 Aturan yang dihasilkan di langkah ke-5 disederhanakan bentuknya supaya mudah dimengerti. Hasilnya dapat dilihat di bawah ini :

```

IF IPKTingkat1 <= 1.9 THEN grade=" gradempat"

IF IPKTingkat1 > 1.9 AND IPKTingkat1 <=2.59 THEN
grade=" gradetiga"

IF IPKTingkat1 > 2.59 AND IPKTingkat1 <= 2.76 THEN
grade=" gradedua"

IF IPKTingkat1 > 2.76 AND IPKTingkat1 <= 3.42 THEN
grade=" gradedua"

IF IPKTingkat1 > 3.42 THEN grade=" gradesatu"
  
```

B. Eksperimen ke-2

Pada eksperimen ke-2, langkah pengumpulan data, integrasi data, dan langkah pembersihan data sama dengan eksperimen ke-1. Data yang digunakan dalam eksperimen ke-2 mempunyai struktur tabel di tabel 1. Di eksperimen 2 ini tidak dilakukan transformasi data. Pada seleksi data dilakukan pemilihan dua atribut, yaitu atribut IPKTingkat 1 dan atribut IPK. Atribut tabel hasil seleksi data ada di tabel III.

TABEL III. ATRIBUT TABEL

Nama	Deskripsi	Type data
IPKTingkat1	IPK setelah menempuh semester 1 dan semester 2	Numerik
IPK	IPK setelah lulus	Numerik

Data mining menggunakan algoritma regresi linear, membentuk aturan:

$$IPK = 0.9919 * IPKTingkat1 + 0.0906$$

Untuk mengevaluasi keakuratan model prediksi digunakan teknik *10-fold cross-validation*.

V. PEMBAHASAN HASIL

A. Pembahasan Hasil Eksperimen 1

Berdasarkan hasil eksperimen tentang klasifikasi didapatkan aturan :

```
IF IPKTingkat1 <= 1.9 THEN grade=" gradempat"
```

Gradempat artinya status berhenti. Artinya jika IPKTingkat1 kurang atau sama dengan 1.9 maka mahasiswa akan berhenti kuliah. Jadi kita bisa memprediksi kelangsungan kuliah mahasiswa berdasarkan IPK tingkat 1.

IPKTingkat1 bisa digunakan untuk memprediksi keberhasilan kuliah mahasiswa. Secara umum, jika IPKTingkat1 rendah, maka IPK terakhir juga rendah dan sebaliknya. Penelitian ini mendapatkan pengetahuan baru, bahwa IPKTingkat1 yang berada di range 2.59 < IPKTingkat1 <= 2.76 (yang berdasarkan peraturan akademik berada di grade 3), ternyata berdasarkan hasil data mining mahasiswa dengan IPK tingkat 1 yang berada di *range* tersebut mendapatkan IPK terakhir di grade yang lebih baik, yaitu di grade 2. Kesimpulan tersebut didukung data perbandingan grade IPK mahasiswa di tabel IV.

TABEL IV. PERBANDINGAN GRADE IPK MAHASISWA

No	IPKTingkat1	Grade berdasarkan peraturan akademik	Grade IPK hasil klasifikasi
1	IPKTingkat1 <=1.9	4	4
2	1.9 < IPKTingkat1 <= 2.59	3	3
3	2.59 < IPKTingkat1 <= 2.76	3	2
4	2.76 < IPKTingkat1 <= 3.42	2	2
5	IPKTingkat1 >3.42	1	1

Pasal 22 ayat 4 peraturan akademik Politeknik Negeri Batam tersebut belum efektif untuk digunakan sebagai dasar hukum penghentian dini mahasiswa yang tidak berpotensi menyelesaikan pendidikan, karena menggunakan batas IPK Tingkat 1 sebesar 1.5. Batas yang seharusnya digunakan adalah 1.9.

B. Pembahasan Hasil Eksperimen 2

Politeknik Negeri Batam menetapkan syarat-syarat yudisium atau kelulusan sebagai berikut : (1) telah dinyatakan lulus dari semua mata kuliah yang disyaratkan oleh kurikulum, (2) IPK minimal 2.0.

Berdasarkan syarat yudisium dan pasal 22 ayat 4, dilakukan prediksi IPK berdasarkan aturan yang dihasilkan dari eksperimen ke-2. Hasil prediksi IPK ada di tabel V. Hasil eksperimen ke-1 menyatakan : IF IPKTingkat1 <= 1.9 THEN grade=" gradempat" (gradempat artinya status

berhenti). Sedangkan berdasarkan hasil eksperimen ke-2, di tabel V baris ke-3, mahasiswa dengan IPKTingkat1 sebesar 1.9 akan memperoleh IPK 1.98. Mahasiswa dengan IPK 1.98 tidak bisa memenuhi syarat minimal yudisium. Hasil kedua eksperimen tersebut sesuai dan tidak saling bertolak belakang. Sedangkan mahasiswa dengan IPKTingkat1 lebih besar dari 1.9 akan mendapat IPK di atas 2.

TABEL V DAFTAR PENGECEKAN IPK

No	IPK Tingkat1	IPK
1	1.49	1.57
2	1.5	1.58
3	1.9	1.98
4	1.99	2.06
5	2	2.07

VI. KESIMPULAN DAN SARAN

Kesimpulan

IPK tingkat 1 bisa digunakan untuk memprediksi kelangsungan dan keberhasilan kuliah mahasiswa dan memprediksi IPK saat lulus. Pasal 22 ayat 4 peraturan akademik Politeknik Negeri Batam harus disempurnakan supaya bisa digunakan secara efektif sebagai dasar penghentian dini mahasiswa yang tidak berpotensi dalam menyelesaikan pendidikan, karena di pasal tersebut batas penghentian dini mahasiswa berdasarkan IPK Tingkat 1 adalah 1.5, sedangkan menurut hasil penelitian ini seharusnya menggunakan batas IPK 1.9.

DAFTAR REFERENSI

- [1] Pedoman Pembelajaran Mahasiswa Politeknik Negeri Batam, 2012.
- [2] J. Bayer, H. Bydzovska, J. Geryk and T. O. L. Popelinsky, "Predicting drop-out from social behaviour of students," in *Proceedings of the fifth International Conference on Educational Data Mining*, Chania, 2012.
- [3] D. GarciaSaiz and M. Zorrilla, "A promising classification method for predicting distance students' performance," in *Proceedings of the fifth International Conference on Educational Data Mining*, Chania, 2012.
- [4] Akcapinar, Cogun and Altun, "Prediction of Perceived Disorientation in Online Learning Environment with Random Forest Regression," in *Proceedings of The Fourth International Conference on Educational Data Mining 2011*, Eindhoven, 2011.
- [5] C. Marquez-Vera, C. Romero and S. Ventura, "Predicting School Failure Using Data Mining," in *Proceedings of The Fourth International Conference on Educational Data Mining 2011*, Eindhoven, 2011.
- [6] J. Zimmermann, K. Brodersen, J. Pellet, E. August and J. Buhmann, "Predicting graduate-level performance from undergraduate achievements," in *Proceedings of The Fourth International Conference on Educational Data Mining 2011*, Eindhoven, 2011.
- [7] G. W. Dekker, M. Pechenizkiy and M. J. Vleeshouwers, "Predicting Students Drop Out: A Case Study," in *Educational Data Mining 2009*, Cordoba, 2009.
- [8] A. Vihavainen, M. Luukkainen and J. Kurhila, "Using Students' Programming Behavior to Predict Success in an Introductory Mathematics Course," in *Proceedings of The Fourth International Conference on Educational Data Mining 2011*, Eindhoven, 2011.
- [9] M. López, J. Luna, C. Romero and S. Ventura, "Classification via clustering for predicting final marks based on student participation in

forums," in *Proceedings of the fifth International Conference on Educational Data Mining*, Chania, 2012.

- [10] S. Ritter, A. Joshi, S. Fancsali and T. Nixon, "Predicting Standardized Test Scores from Cognitive Tutor Interactions," in *Proceedings of the 6th International Conference on Educational Data Mining(EDM 2013)*, Memphis, 2013.
- [11] K. Kelly, I. Arroyo and N. He_ernan, "Using ITS Generated Data to Predict Standardized Test Scores," in *Proceedings of the 6th International Conference on Educational Data Mining*, Memphis, 2013.
- [12] J. Han, *Data Mining : Concepts and techniques*, Morgan Kaufmann, 2006.
- [13] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Technique*, Morgan Kauffman, 2005.