

# PENGLASIFIKASIAN PRESTASI PELAJAR MENGGUNAKAN PENDEKATAN PEMBELAJARAN MESIN

Khairil Anuar Bin Azman Hisham

Hazura Mohamed

Nor Samsiah Sani

*Fakulti Teknologi dan Sains Maklumat, Universiti Kebangsaan Malaysia*

## ABSTRAK

Serasi dengan pembangunan teknologi, proses perlombongan data dianggap sebagai sebuah anjakan paradigma, teknologi ini telah berjaya diterapkan ke pelbagai bidang termasuk bidang pendidikan. Ramalan dan analisis prestasi akademik pelajar adalah penting untuk kemajuan akademik pelajar namun ianya adalah tugas yang mencabar kerana terdapat pelbagai faktor yang berbeza yang mempengaruhi prestasi pelajar. Disamping bagaimana bagi mengklasifikasi prestasi pelajar sama ada cemerlang atau sederhana. Justeru, kajian ini bertujuan mengklasifikasi prestasi pelajar Fakulti Teknologi dan Sains Maklumat (FTSM) menggunakan kaedah pembelajaran mesin. Beberapa model pembelajara mesin iaitu Naïve Bayes (NB), Decision Tree (DT), k-Neural Network (K-NN) dan Random Forest (RF) diguna bagi mengklasifikasi prestasi pelajar. Daripada empat model yang diadaptasi dalam kajian ini, hasil menunjukkan DT ialah model terbaik dalam mengklasifikasi prestasi pelajar.

Katakunci: Perlombongan data, pembelajaran mesin, klasifikasi prestasi

## 1.0 Pengenalan

Pendidikan adalah kunci utama dalam pembangunan dalam progress ekonomi Prestasi akademik pelajar dipengaruhi oleh pelbagai faktor seperti latar belakang keluarga, sosio-ekonomi, minat, serta bimbingan rakan sebaya. (Baradwaj & Pal 2014). Bagi para saintis mahupun tenaga pengajar, kebolehan untuk memperoleh pengetahuan mengenai kaitan kepada faktor tersebut adalah suatu yang istimewa. Ianya kerana, para saintis dan ilmuwan

mampu untuk merancang struktur kurikulum yang baik dan bermanfaat untuk menggilap potensi para pelajar melalui perancangan akademik yang baik. Begitulah hakikat konsep data sains dalam proses perlombongan data prestasi pelajar melalui gred kumulatif. Dewasa ini, proses aktiviti perlombongan data di dalam bidang pendidikan semakin meningkat. Hal ini kerana kementerian pengajian amatlah menitikberatkan dalam menghasilkan jumlah graduan yang cemerlang dari segi akademik. proses perlombongan data pendidikan ataupun dalam bahasa inggerisnya, *Educational data Mining* (EDM) merujuk kepada teknik, peralatan dan penyelidikan yang direka untuk memperoleh data secara automatik daripada repositori data yang diperoleh daripada sistem maklumat akademik pelajar (Nithya et al. 2016).

Teknik EDM ini merupakan salah satu data yang sukar kerana jumlah data yang diperoleh dari pangkalan data adalah sangat banyak dan berterusan sepanjang masa. Hasil daripada ramalan prestasi pelajar adalah sangat penting dalam bidang perguruan dan kebolehan untuk meningkatkan prestasi pelajar adalah matlamat utama para ilmuwan dan pengkaji. Sekiranya institusi akademik mampu untuk meramal prestasi akademik pelajar ketika sebelum mereka menduduki peperiksaan akhir, maka tenaga pengajar mampu untuk berusaha dengan cara yang berbeza untuk menyusun strategi bagi meningkatkan dan membantu pelajar yang berprestasi rendah untuk berjaya (Sarker et al. 2012). Malahan, daripada hasil perlombongan data, para ilmuwan mampu untuk mengenal pasti atribut yang berkemungkinan mempengaruhi peratusan kejayaan sesebuah subjek kursus.

Sektor pendidikan di Malaysia terutamanya sektor pengajian tinggi di Malaysia adalah bersifat kompetitif. Cabaran utama dalam universiti moden adalah untuk menganalisis prestasi graduan yang dihasilkan pada sesi terkini dan mengenal pasti keunikan struktur kurikulum pada sesi tersebut untuk dijadikan sebagai rujukan kepada penambahbaikan struktur kurikulum untuk sesi yang akan datang. Objektif perlombongan data pelajar yang terpilih untuk memasuki universiti ini adalah supaya pihak universiti mampu untuk merancang kempen kemasukan pelajar dan mendekati pelajar yang mempunyai potensi dan mampu untuk memberi impak yang baik untuk universiti pada masa akan datang.

Serasi dengan pembangunan teknologi, proses perlombongan data dianggap sebagai sebuah anjakan paradigma, teknologi ini telah berjaya diterapkan ke pelbagai bidang

termasuk pendidikan. Baru-baru ini, minat minat penyelidikan semakin meningkat dalam menggunakan perlombongan data dalam pendidikan. bidang baharu ini yang dipanggil perlombongan data pendidikan (EDM) (Mueen 2016). Objektif EDM adalah untuk membangunkan kaedah baharu untuk meneroka data pendidikan untuk menentukan kegunaan sistem pembelajaran dan analisis prestasi akademik pelajar. Ramalan dan analisis prestasi akademik pelajar adalah penting untuk kemajuan akademik pelajar namun ianya adalah tugas yang mencabar kerana pengaruh faktor-faktor yang berbeza yang mempengaruhi prestasi pelajar seperti faktor keluarga, profil psikologi, pencapaian akademik terdahulu, dan interaksi pelajar dengan rakan sekelas beserta tenaga pengajar (Kotsiantis 2012). Amra (2017) telah menjalankan kajian prestasi pelajar di universiti dengan menggunakan teknik pembelajaran mesin dan antara faktor yang dilihat ialah faktor sosioekonomi pelajar seperti jantina, fakulti dan status perkahwinan.

Justeru dengan menggunakan pendekatan yang sama iaitu kaedah pembelajaran mesin, prestasi pelajar Fakulti Teknologi dan Sains Maklumat (FTSM) dianalisis berdasarkan beberapa model pembelajaran mesin bagi menentukan model yang sesuai dalam pengklasifikasi prestasi pelajar.

### **Pembelajaran Mesin**

Mesin pembelajaran merangkumi beberapa fasa iaitu fasa pembersihan data, integrasi data dan pembangunan model. Penumpuan terhadap komputasi dan komunikasi telah menghasilkan sebuah masyarakat yang hidup dengan suapan data dan menghasilkan data dengan tidak sengaja pada setiap masa seperti data sentimen melalui penggunaan media sosial, ataupun pengisian informasi data melalui borang digital yang akhirnya membawa kepada simpanan ke dalam pangkalan data. Namun, kebanyakan informasi yang terhasil adalah bersifat mentah dan berpotensi untuk diekstrak dan dianalisis untuk menjana pengetahuan penting.

Mesin pembelajaran menyediakan kemudahan asas teknikal kepada perlombongan data. Mesin pembelajaran digunakan untuk mengekstrak informasi daripada data di mana informasi tersebut boleh dinilai dalam bentuk yang boleh difahami dan digunakan untuk pelbagai tujuan. Seperti mana – mana teknologi yang berkembang yang menikmati perhatian komersil yang sengit, aplikasi perlombongan data sering digunakan atas alasan hasil analisis

yang terhasil senang untuk digembar – gemburkan lebih – lebih lagi di dalam sidang akhbar. Namun begitu, tiada yang istimewa di sebalik pembelajaran mesin kecuali kaedah statistik yang menggunakan pelbagai jenis algoritma, peraturan dan teknik yang mana ianya lebih canggih berbanding kaedah statistik lama. Seperti pasaran telefon bimbit yang berkembang dengan cepat, teknologi menafsir data juga berkembang dari teknik statistik lama kepada teknik statistik yang lebih maju.

Antara teknik yang sering digunakan di dalam perlombongan data adalah teknik klasifikasi. Di dalam teknik klasifikasi, sesebuah model dibina daripada contoh set latihan untuk menetapkan ketepatan label atau kelas kepada set data ujian. Terdapat dua bahagian di dalam teknik klasifikasi iaitu bahagian latihan dan bahagian ujian. Di dalam fasa latihan, sebuah model dibina dengan menggunakan data yang dikenali sebagai set latihan yang mana model tersebut memahami dan mengenal pasti setiap atribut dan juga rekod beserta kelas label. Selepas penjanaan model, ianya sekali lagi digunakan tetapi pada kali ini, ianya digunakan untuk mengenal pasti kelas label daripada set data ujian di mana nilai setiap rekod di atribut tidak dikenali. Terdapat beberapa teknik mesin pembelajaran seperti *decision tree* (DT), rangkaian neural, *naive bayes* (NB), k- jiran terdekat (k-NN) dan juga *support vector machine* (SVM) (Charu 2014).

Algoritma Naïve Bayes merupakan sebuah teknik klasifikasi menggunakan gabungan kebarangkalian dan statistik yg dikemukakan oleh tokoh bernama Thomas Bayes (Marlina et al. 2016). Algoritma Naive Bayes merupakan algoritma untuk data yang bersifat supervised di mana, fitur – fitur daripada data adalah bebas dan tidak terikat dengan fitur lain secara statistik. Teori ini bergantung kepada algoritma dengan membuat andaian bahawa pembolehubah di dalam data juga tidak bergantung kepada fitur – fitur, yang mana secara amnya algoritma ini bersifat naif. (Sebagai contoh, tiada cara untuk mesin / algoritma Naïve Bayes ini untuk mengetahui tentang sesuatu pembolehubah sekiranya diletakkan pembolehubah baru. Meskipun algoritma Naïve Bayes bersifat naif dan hanya dilakukan untuk melakukan andaian terhadap data, algoritma ini masih digunakan sehingga sekarang dan menunjukkan bahawa ianya adalah salah satu teknik mesin pembelajaran yang terkenal.

Bhardwaj dan Pal (2014) melakukan kajian perlombongan data menggunakan teknik klasifikasi Naive bayes terhadap pelajar sarjana muda. Kajian tersebut mendapati bahawa faktor yang paling mempengaruhi prestasi pelajar ketika di peringkat pelajaran tinggi adalah

daripada alam sekolah menengah lagi, yang mana diperacyai bahawa pelajar yang mempunyai prestasi yang baik ketika di alam sekolah menengah mampu untuk mencapai keputusan yang baik ketika menduduki peperiksaan di peringkat pengajian tinggi (Baradwaj & Pal 2014).

Pokok keputusan (DT) ataupun dalam kamus Bahasa Inggeris ialah *Decision Tree* (DT), merupakan sebuah teknik pembelajaran mesin yang sangat popular dalam bidang pembelajaran mesin di mana ianya hasil daripada proses pemilihan atribut paling penting untuk membezakan sifat antara label kelas. Algoritma tersebut mampu untuk memperoleh markah kejituan yang sangat tinggi dan sangat versatil untuk digunakan dalam pelbagai bentuk data. Malahan hasil daripada teknik DT ini juga senang untuk ditafsir oleh para pengguna mesin pembelajaran dengan penjaanan petua sebagai tafsiran. Salah satu contoh penggunaan DT di dalam bidang perubatan adalah di dalam koteks diagnosis penyakit melalui gejala penyakit, di mana kelas yang ditakrifkan oleh pokok keputusan boleh sama ada jenis klinikal yang berbeza atau keadaan, atau pesakit dengan keadaan yang harus menerima terapi yang berlainan (Nithya *et al.* 2016).

Teknik mesin pembelajaran pokok rawak (RF) adalah merupakan teknik yang menggunakan konsep pembungkusan di mana model ini akan menghasilkan beberapa pokok keputusan dan hasil ketepatan paling tinggi di dalam model tersebut akan dipilih untuk menggambarkan ketepatan model. (Abubakar & Ahmad 2017). Berbanding DT, RF memecahkan setiap rantai dengan menggunakan subset daripada algoritma peramal untuk mendapatkan keputusan yang baik dan teknik ini sangat selesa untuk digunakan terhadap data yang berlebihan.

Di dalam artikel (Hussain & Muhsen 2019) beliau menggunakan beberapa model klasifikasi seperti DT, RF, dan NB. Di dalam kajian literatur tersebut beliau menggunakan beberapa atribut di dalam set data seperti faktor sosioekonomi pelajar dan juga prestasi akademik pelajar seperti peratusan kehadiran, dan waktu mengulangkaji. Di dalam jurnal tersebut, terdapat perbezaan yang signifikan di model RF apabila melalui pemilihan atribut berbanding dengan atribut yang asal

Analisis secara kluster telah digunakan secara meluas di dalam pelbagai aplikasi seperti penyelidikan di pasaran dan analisis corak sentimen (Dey n.d.). Di dalam bidang bisnes, teknik kluster dapat membantu pemasar untuk menemui minat para pelanggan. Algoritma k-

NN adalah berdasarkan prinsip di mana jika sampel adalah hampir sama maka ianya berada dalam satu kelompok yang sama (Jadhav & Channe 2016). Untuk algoritma ini, jiran yang terdekat dikira berdasarkan nilai  $k$  iaitu nilai yang menentukan berapa jumlah jiran untuk diperiksa sebelum penyediaan sesebuah kluster (Amra 2017).

### **Kaedah Kajian**

Kajian ini mempunyai empat fasa dalam pengelasan penentuan keputusan prestasi pelajar iaitu bermula daripada proses penyediaan data, pembersihan data, integrasi data dan diakhiri dengan proses pembangunan model pengelasan.

Proses penyediaan data merupakan peringkat pertama dalam amalan teknik perlombongan data. Dalam kajian ini, proses penyediaan data yang dilakukan ialah proses integrasi data dan pembersihan data. Kedua proses ini wajib bagi memastikan set data lengkap boleh diguna pakai untuk tujuan pemodelan di mana ianya mampu untuk menghasilkan model pengelasan yang betul. Selain itu, dengan proses pembersihan data juga mampu untuk menurunkan peratusan ralat daripada model yang dicipta. Data yang diperoleh adalah data prestasi akademik pelajar Fakulti Teknologi Sains Maklumat (FTSM). Oleh itu, proses penyediaan data perlu dilakukan bagi menyediakan set data lengkap untuk tujuan permodelan. Proses – proses utama yang dilakukan ialah proses integrasi data, dan proses pembersihan data.

Di dalam kajian ini, model ujian pengelasan yang dipilih untuk digunakan ialah model pengelasan seperti model pokok keputusan (DT), model pokok keputusan rawak (RT), model Naive Bayes (NB), dan  $k$ -jiran terdekat ( $k$ -NN). Alasan kepada pemilihan model di atas adalah kerana label kelas untuk projek kali ini adalah dalam bentuk polinomial. Oleh itu, penggunaan model SVM tidak dapat digunakan untuk set data ini kerana model SVM hanyalah eksklusif kepada set data di dalam bentuk binomial.

Selain itu, teknik pemecahan data yang digunakan ialah teknik *cross-fold validation*. Dalam teknik *cross fold validation*, nisbah 4, 6, 8 dan 10 dengan cara pembahagian berstrata dan rawak. Hasil dapatan analisis prestasi model seperti ketepatan, kejituan, dapatan semula dan pengiraan-F dijadualkan untuk dibuat perbandingan. Model yang mendapat keputusan ramalan dan ketepatan yang tinggi bermaksud bahawa model tersebut mampu melakukan

tugas ramalan terbaik untuk sesuatu jenis data. Oleh sebab itu, kadar peratusan ketepatan adalah nilai utama untuk diperhatikan di dalam soal pemilihan model yang terbaik.

$$\text{Ketepatan} = \frac{K1+K2+K3}{n} \quad - (1)$$

$$\text{kejituan} = \frac{K1}{K1 + xK1} \quad - (2)$$

## Hasil Kajian dan Perbincangan

Jadual 1 hingga Jadual 4 menunjukkan pengelasan tunggal yang dilakukan, model DT menghasilkan ketepatan paling tinggi iaitu dengan purata 96.16% untuk jenis lipatan strata berbanding model NB dan model RF iaitu dengan purata 96.04% untuk model NB dan 94.54% untuk model RF dan akhir sekali diikuti dengan model k-NN iaitu dengan purata sebanyak 80.35% untuk jenis lipatan strata. Dengan mempunyai saiz data yang kecil iaitu sebanyak 274 data, model DT mempunyai kelebihan untuk memberi keputusan yang baik dan keputusan yang terbaik adalah pada pecahan 8 kali lipatan dan 10 kali jenis lipatan strata. Bagi jenis lipatan dalam bentuk rawak pula, model yang mempunyai nilai ketepatan yang tinggi adalah model NB dengan purata 95.74%, diikuti dengan model DT, RF dan k-NN dengan masing – masing bernilai 95.44, 93.13 dan 79.27 untuk model jenis lipatan rawak. Jika dilihat pada Jadual 2, korelasi pada bilangan lipatan dan peratusan pada model DT adalah berkadar terus di mana semakin besar bilangan lipatan, semakin tinggi nilai ketepatan dan nilai ketepatan dalam kadar tepu bermula dari bilangan lipatan lapan.

Jadual 1 Keputusan parameter ketepatan model pengelasan NB melalui pembolehubah jenis lipatan dan bilangan lipatan

Model	Jenis Lipatan	Bil. Lipatan	Masa	Pengelasan Betul	Pengelasan Salah	Ketepatan (%)
NB	Strata	4	0.1	238	9	96.35
		6	0.1	238	9	96.34
		8	0.1	236	11	95.55
		10	0.1	237	10	95.95
	Rawak	4	0.1	237	10	95.93
		6	0.1	238	9	96.35
		8	0.1	235	12	95.13
		10	0.2	236	11	95.55

Jadual 2 Keputusan parameter ketepatan model pengelasan DT melalui pembolehubah jenis lipatan dan bilangan lipatan

<b>Model</b>	<b>Jenis Lipatan</b>	<b>Bil. Lipatan</b>	<b>Masa</b>	<b>Pengelasan Betul</b>	<b>Pengelasan Salah</b>	<b>Ketepatan (%)</b>
<i>DT</i>	Strata	4	0.1	235	12	95.15
		6	0.1	237	10	95.99
		8	0.1	240	7	96.75
		10	0.1	240	7	96.75
	Rawak	4	0.1	233	14	93.88
		6	0.1	240	7	96.76
		8	0.1	234	13	94.34
		10	0.2	240	7	96.78

Jadual 3 Keputusan parameter ketepatan model pengelasan RF melalui pembolehubah jenis lipatan dan bilangan lipatan

<b>Model</b>	<b>Jenis Lipatan</b>	<b>Bil. Lipatan</b>	<b>Masa</b>	<b>Pengelasan Betul</b>	<b>Pengelasan Salah</b>	<b>Ketepatan (%)</b>
<i>RF</i>	Strata	4	0.1	233	14	94.34
		6	0.1	237	10	95.94
		8	0.1	230	17	93.12
		10	0.1	234	13	94.77
	Rawak	4	0.1	224	23	90.66
		6	0.1	234	13	94.34
		8	0.1	233	14	93.53
		10	0.2	232	15	93.98

Jadual 4 Keputusan parameter ketepatan model pengelasan k-NN melalui pembolehubah jenis lipatan dan bilangan lipatan

Model	Jenis Lipatan	Bil. Lipatan	Masa	Pengelasan Betul	Pengelasan Salah	Ketepatan (%)
<i>k-NN</i>	Strata	4	0.1	203	44	81.79
		6	0.1	200	48	80.55
		8	0.1	200	48	80.55
		10	0.1	194	53	78.52
	Rawak	4	0.1	199	48	79.36
		6	0.1	200	47	79.74
		8	0.1	195	52	78.98
		10	0.2	199	48	79.00

Bagi memalar dan mengadili keputusan untuk pemilihan model yang terbaik, markah kejitian model dalam kategori bilangan lipatan 10 secara strata digunakan. Alasannya adalah kerana pecahan dalam bentuk strata adalah lebih stabil berbanding dengan pecahan berbentuk rawak di mana kemungkinan besar set pecahan menjadi tidak seimbang di dalam pecahan secara rawak terutama di dalam set data yang mana nisbah perbezaan di antara kelas label adalah sangat ketara. Jadual 5 menunjukkan keputusan model terbaik untuk setiap model pengelasan dan kriteria yang dinilai.

Jadual 5 Keputusan keseluruhan prestasi model terbaik model pengelasan

Model	NB	DT	RF	k-NN
Ketepatan	95.95%	<b>96.75%</b>	94.77%	78.52%
Kejitian	<b>1</b>	0.95	<b>1</b>	0.5
Dapatan semula	0.68	<b>1</b>	0.84	0.1
Pengiraan – F	0.809	<b>0.974</b>	0.913	0.166
Sisihan piawai	5.2	<b>3.22</b>	6.09	4.31

Selain itu, parameter kejitian, dapatan semula dan pengiraan-F juga diperhati dalam melihat prestasi model pengelasan yang dipersembahkan di dalam Jadual 5. Di dalam model penentuan keputusan gred kejayaan pelajar, model NB dan RF mendapat kadar kejitian yang tinggi iaitu niainya 1 berbanding model-model lain. Kadar kejitian yang tinggi ini bermaksud model NB dan RF berupaya untuk mengelaskan gred pelajar dengan lebih baik berbanding dengan algoritma lain. Selain itu, sekiranya jumlah kelas label pilihan iaitu kelas pertama

adalah seimbang dan banyak, kadar kejituan model pengelas juga dapat ditingkatkan dan mampu mengurangkan ralat di dapatan semula.

Hal ini kerana dapatan semula juga berupaya menunjukkan sama ada model pengelas yang dibangunkan adalah baik atau tidak untuk mengelaskan pelajar mendapat gred kelas pertama atau tidak. Jika dilihat pada Jadual 5, model DT mempunyai nilai dapatan semula yang tinggi iaitu 1, diikuti dengan RF sebanyak 0.84, NB sebanyak 0.68 dan k-NN sebanyak 0.1. Oleh itu, model DT mampu untuk mengelaskan pelajar yang mendapat gred kelas pertama daripada gred kelas kedua tinggi dan rendah. Meskipun nilai kejituan DT hanyalah 0.95, namun nilai itu hampir dekat dengan nilai yang dihasilkan oleh model NB dan RF. Kemungkinannya dengan model DT, peraturan yang dijana oleh pokok keputusan mampu untuk menggunakan atribut yang tersedia dalam pengelasan prestasi pelajar. Meskipun set kelas label adalah tidak seimbang iaitu kelas label sasaran adalah 20% dari total set data, namun model pengelasan DT masih mampu untuk melaksanakan tugas dengan baik.

Pengiraan-F mempunyai maksud yang intuitif. Ianya mampu untuk memberitahu kita ketepatan model klasifikasi dan juga betapa versatil model klasifikasi itu. Dengan kadar kejituan yang tinggi namun dapatan semula yang rendah, model klasifikasi akan mendapat nilai ketepatan yang tinggi namun ianya mampu untuk terlepas beberapa data ataupun atribut yang sukar untuk diklasifikasikan. Dalam projek ini, nilai pengiraan-F yang paling tinggi adalah DT iaitu sebanyak 0.974 diikuti dengan RF sebanyak 0.913 dan NB sebanyak 0.809. Model k-NN mempunyai nilai pengiraan-F yang paling sedikit iaitu sebanyak 0.166. Selain itu, sisihan piawai merupakan salah satu nilai penting yang patut dilihat ketika pembentukan model pengelasan untuk sesuatu data. Hal ini kerana sifat sisihan piawai adalah untuk menunjukkan betapa stabil sesebuah model itu dan ianya berkadar songsang dengan kestabilan data. Semakin rendah sisihan piawai, semakin stabil model itu. Model yang mempunyai sisihan piawai paling rendah berbanding 3 model lain adalah model DT dengan nilai sebanyak 3.22. Meskipun model DT mempunyai nilai kejituan yang rendah berbanding model lain namun model DT menunjukkan prestasi yang baik dalam setiap parameter yang berlainan. Nilai tertinggi untuk setiap parameter mempunyai tulisan 'bold' untuk mengenal pasti parameter terbaik.

## **Kesimpulan**

Kajian ini berjaya mengenal pasti teknik dan algoritma bagi mengelas prestasi pelajar dengan mencadangkan model DT ialah model terbaik dalam pengelasan prestasi pelajar. Model ini berjaya meningkatkan keupayaan bagi menentukan pengelasan prestasi pelajar dan berkemungkinan memberi manfaat kepada kajian ilmu berkaitan bidang pembelajaran mesin dan pemerolehan maklumat akan datang. Kajian yang dijalankan dan penemuan yang diperolehi, dapat memberi panduan dan asas kepada kajian pengklasifikasian akan datang terutama dalam bidang pendidikan. Hasil penemuan model pengelasan dan pemerolehan maklumat ini juga membantu mengukuhkan lagi hujah kajian ilmiah yang lepas dan berpotensi untuk memberi kesan positif kepada proses pembuatan model klasifikasi pelajar.

**RUJUKAN**

- Abubakar & Ahmad. 2017. Prediction of Students' Performance in ELearning Environment Using Random Forest. *IJIC* Vol. 7:2 (2017) 1-5
- Academy, B., Sciences, O. F. & Technologies, I. 2013. Predicting Student Performance by Using Data Mining Methods for Classification Dorina Kabakchieva 13(1): 61–72. doi:10.2478/cait-2013-0006
- Amra, I. A. A. 2017. Students performance prediction using KNN and Naïve Bayesian (December 2018). doi:10.1109/ICITECH.2017.8079967
- Baradwaj, B. K. & Pal, S. 2014. Mining Educational Data to Analyze Students " Performance (October 2011). doi:10.14569/IJACSA.2011.020609
- Buenaño-fern, D. & Gil, D. 2019. Application of Machine Learning in Predicting Performance for Computer Engineering Students : A Case Study 1–18.
- Cao, Y., Gao, J., Lian, D., Rong, Z., Shi, J., Wang, Q., Wu, Y., et al. 2018. Orderliness predicts academic performance : behavioural analysis on campus lifestyle.
- Charu, C. A. "An Introduction to Data Classification" Data Classification, Chapman and Hall/CRC, pp. 1-36, 2014..
- Cortez, P., Silva, A., Trees, D. & Forest, R. 2003. USING DATA MINING TO PREDICT SECONDARY SCHOOL STUDENT PERFORMANCE 2003(2000).
- Dey, L. (n.d.). Sentiment Analysis of Review Datasets using Naïve Bayes ' and K -NN Classifier.
- H. Bhavsar and A. Ganatra, "A Comparative Study of Training Algorithms for Supervised Machine Learning", *International Journal of Soft Computing and Engineering (IJSCE)*, Vol. 2, Issue. 4, September 2012

- Hashemi, M. & Hall, M. 2018. Visualization , Feature Selection , Machine Learning : Identifying the Responsible Group for Extreme Acts of Violence. *IEEE Access* 6: 70164–70171. doi:10.1109/ACCESS.2018.2879056
- Hussain, S. & Muhsen, Z. F. 2019. Prediction Model on Student Performance based on Internal Assessment using Deep Learning Prediction Model on Student Performance based on Internal Assessment using Deep Learning (May). doi:10.3991/ijet.v14i08.10001
- Jadhav, S. D. & Channe, H. P. 2016. Comparative Study of K-NN , Naive Bayes and Decision Tree Classification Techniques 5(1): 2014–2017.
- Kabakchieva D (2012) Student performance prediction by using data mining classification algorithms. *IJCSMR* 1: 686-690
- Khasanah, Annisa & Harwati, Harwati. (2017). A Comparative Study to Predict Student's Performance Using Educational Data Mining Techniques. *IOP Conference Series: Materials Science and Engineering*. 215. 012036. 10.1088/1757-899X/215/1/012036
- Kotsiantis, S. B. "Use of machine learning techniques for educational proposes: A decision support system for forecasting students grades," *Artificial Intelligence Review*, 37(4), pp.331–344, 2012,
- Marlina, L., Putera, A. & Siahaan, U. 2016. Data Mining Classification Comparison ( Naïve Bayes and C4 . 5 Algorithms ) 38(7): 380–383.
- Mueen, A. 2016. Modeling and Predicting Students ' Academic Performance Using Data Mining Techniques (November): 36–42. doi:10.5815/ijmecs.2016.11.05
- Nithya, P., Umamaheswari, B. & Umadevi, A. 2016. A Survey on Educational Data Mining in Field of Education 5(1): 69–78.
- Nyamapfene, A. 2017. Does class attendance still matter ? (June 2010). doi:10.11120/ened.2010.05010064

Petri, C. (n.d.). Decision Trees.

Reis, I., Baron, D. & Shahaf, S. 2019. Probabilistic Random Forest : A machine learning algorithm for noisy datasets Probabilistic Random Forest (September).

Saa, A. A. 2016. Educational Data Mining & Students ' Performance Prediction 7(5): 212–220.

Sarker, F., Tiropanis, T. & Davis, H. C. 2012. Students ' Performance Prediction by Using Institutional Internal and External Open Data Sources.

Shields, L.; Newman, A.; Satz, D. Equality of Educational Opportunity. In Stanford Encyclopedia of Philosophy; Zalta, E., Ed.; Stanford University: Stanford, CA, USA, 2017

Song, Y. & Lu, Y. 2015. Decision tree methods : applications for classification and prediction 27(2): 130–135.

Stumm, S. Von, Hell, B. & Chamorro-premuzic, T. 2015. The Hungry Mind : Intellectual Curiosity Is the Third Pillar of Academic Performance. doi:10.1177/1745691611421204

Zah, W. A. N., Ali, W. A. N., Hamzah, R., Abu, R., Tarmizi, R. A., Bakar, K. A. B. U. & Ismail, H. 2008. Kemahiran pengurusan masa pelajar universiti tempatan: status dan hubungannya dengan pencapaian 49: 49–64.