

**PEMILIHAN FITUR YANG MEMPENGARUHI
PRESTASI AKREDITASI MENGGUNAKAN
KAEDAH PEMBELAJARAN MESIN**

NOR KHAMSATUN BINTI SUBOH

UNIVERSITI KEBANGSAAN MALAYSIA

PEMILIHAN FITUR YANG MEMPENGARUHI PRESTASI AKREDITASI
MENGUNAKAN KAEDAH PEMBELAJARAN MESIN

NOR KHAMSATUN BINTI SUBOH

PROJEK YANG DIKEMUKAKAN UNTUK MEMENUHI SEBAHAGIAN
DARIPADA SYARAT MEMPEROLEH IJAZAH
SARJANA SAINS DATA

FAKULTI TEKNOLOGI DAN SAINS MAKLUMAT
UNIVERSITI KEBANGSAAN MALAYSIA
BANGI

2024

PENAKUAN

Saya akui karya ini adalah hasil kerja saya sendiri kecuali nukilan dan ringkasan yang tiap-tiap satunya telah saya jelaskan sumbernya.

26 Jun 2024

NOR KHAMSATUN BINTI SUBOH
P119940

PENGHARGAAN

Alhamdulillah, syukur kehadiran Allah S.W.T kerana dengan limpah kurnia dan izinNya, dapat saya menyiapkan kajian projek akhir pengajian saya dengan jayanya.

Sekalung budi dan setinggi-tinggi penghargaan yang tidak terhingga ingin saya ucapkan kepada YBrs. Prof. Madya Dr. Mohd Zakree Ahmad Nazri selaku penyelia saya yang tidak putus-putus memberi bimbingan dan tunjuk ajar sepanjang saya melaksanakan kajian ini. Tanpa sokongan, doa, dan dorongan beliau, pasti sukar untuk saya menyiapkan kajian ini. Segala ilmu, nasihat dan pandangan yang diberi amat bermakna untuk saya gunakan apabila kembali bekerja kelak. Ucapan terima kasih juga kepada penyelaras program Sarjana Sains Data, Dr. Wandeeep Kaur A/P Ratan Singh dan semua pengajar di FTSM yang tidak kenal erti lelah mendidik, mencurahkan ilmu dan memberi bantuan sepanjang pengajian saya di UKM.

Penghargaan ini juga ditujukan kepada Jabatan Perkhidmatan Awam Malaysia yang membiayai pengajian saya hingga selesai dan kepada Agensi Kelayakan Malaysia (Malaysian Qualifications Agency, MQA) yang memberi peluang kepada saya untuk melanjutkan pengajian ini dan memberi kerjasama dan kebenaran untuk menggunakan data program akreditasi yang digunakan di dalam kajian ini.

Buat keluarga tercinta, khususnya suami dan anak-anak, Pardon bin Munawar, Iman Hakimi bin Pardon, Iman Nasuha binti Pardon dan Iman Nukman bin Pardon, terima kasih yang tidak terhingga di atas kasih sayang, dorongan, kepercayaan, kesabaran dan terus mendoakan kejayaan saya khususnya dalam menyiapkan kajian ini. Ucapan terima kasih juga buat adik-beradik tersayang yang sentiasa memahami dan memberi sokongan sepanjang tempoh pengajian ini.

Buat rakan-rakan seperjuangan bidang Sarjana Sains Data UKM, jutaan terima kasih kerana tidak pernah lokek untuk berkongsi ilmu, pandangan dan memberi bantuan. Akhir sekali, sekalung penghargaan buat semua yang terlibat sama ada secara langsung atau tidak langsung sepanjang perjalanan pengajian sarjana ini.

Semoga Allah S.W.T membalas segala budi baik kalian.

ABSTRAK

Akreditasi program akademik adalah penting sebagai jaminan kualiti bagi program yang ditawarkan oleh Pemberi Pendidikan Tinggi (PPT). Program yang berkualiti dapat melahirkan graduan bermutu dan meningkatkan kebolehpasaran program akademik dan graduan. Proses akreditasi bermula dengan PPT menghantar permohonan berserta dokumen untuk dinilai berasaskan Kerangka Kelayakan Malaysia dan standard yang ditetapkan. Tempoh atau prestasi pasukan yang menilai dokumen berbeza antara satu sama lain kerana dipengaruhi pelbagai faktor. Terdapat faktor yang diketahui dan ada juga faktor lain yang tidak diketahui. Bagi memastikan program ini dapat ditawarkan pada masa yang ditetapkan, pengurusan memerlukan kaedah untuk meramal prestasi penilaian program agar boleh melakukan intervensi dari awal. Oleh yang demikian, kajian ini dibuat untuk mengetahui teknik pemilihan fitur dan fitur optimum yang mempengaruhi prestasi tempoh penilaian program akreditasi baharu sama ada proses penilaian dapat diselesaikan dalam tempoh yang ditetapkan atau di luar tempoh yang ditetapkan. Kajian ini turut membangunkan model pengelasan berprestasi baik dalam meramal prestasi penilaian program akreditasi baharu. Data yang digunakan di dalam kajian ini ialah data penerimaan permohonan penilaian program Akreditasi Provisional dan Akreditasi Penuh bagi tempoh bermula dari tahun 2019 sehingga tahun 2022 yang telah selesai diproses. Kajian ini menggunakan lima (5) teknik pemilihan fitur iaitu penapis *Information Gain* (IG), pembalut pemilihan berjujukan ke depan (PBD), pembalut penyingkiran berjujukan ke belakang (SBB), carian rawak Algoritma Genetik (AG) dan hibrid penapis-pembalut PBD serta tiga (3) algoritma pengelasan iaitu Regresi Logistik (RL), Naive Bayes (NB) dan Hutan Rawak (HR). Dapatan kajian ialah teknik pembalut SBB dengan 12 fitur optimum mempengaruhi prestasi penilaian dan model pengelasan dengan prestasi terbaik iaitu HR dengan skor-f 83.77%. Dengan model yang terpilih, kajian ini boleh membantu agensi untuk penambahbaikan ke atas faktor-faktor yang mempengaruhi proses penilaian seterusnya membuat penambahbaikan bagi meningkatkan prestasi penilaian program akreditasi baharu di dalam tempoh yang ditetapkan.

FEATURE SELECTION FOR IMPROVING ACCREDITATION PERFORMANCE USING MACHINE LEARNING TECHNIQUES

ABSTRACT

Accreditation of academic programs is important as a guarantee of quality for programs offered by Higher Education Providers (HEP). Quality programs can produce quality graduates and increase the marketability of academic programs and graduates. The accreditation process begins with HEP sending an application along with documents to be assessed based on the Malaysian Qualifications Framework and standards. The duration or performance of the team that evaluates the document differs from each other because it is influenced by various factors. There are known factors and there are also other unknown factors. In order to ensure that this program can be offered at the appointed time, management needs a method to predict the performance of the program evaluation so that management can intervene from the beginning. Therefore, this study was made to find out the feature selection technique and optimal features that affect the performance of the evaluation period of the new accreditation program whether the evaluation process can be completed within the specified period or outside the specified period. This study also developed a classification model that performed well in predicting the evaluation performance of new accreditation programs. The data used in this study are the application programs acceptance of the Provisional Accreditation and Full Accreditation from year 2019 until year 2022, which have been processed. This study uses five (5) feature selection techniques, namely filter method Information Gain (IG) technique, three (3) wrapper method - forward selection technique, backward elimination technique and genetic algorithm random search technique and also hybrid method filter-wrapper technique. Three (3) classification algorithms namely Logistic Regression, Naive Bayes and Random Forest are used in this research. The findings showed that the wrapper backward elimination technique with 12 optimal features affects the assessment performance and the classification model with the best performance is Random Forest with an F-score of 83.77%. With the selected model, this study can assist agencies in improving factors that influence the assessment process and make improvements to enhance the performance of the new accreditation program assessment within the specified timeframe.

KANDUNGAN

| | | Halaman |
|--------------------------|---|----------------|
| PENGAKUAN | | ii |
| PENGHARGAAN | | iii |
| ABSTRAK | | iv |
| ABSTRACT | | v |
| KANDUNGAN | | vi |
| SENARAI JADUAL | | ix |
| SENARAI ILUSTRASI | | xi |
| SENARAI SINGKATAN | | xiv |
| | | |
| BAB I | Pengenalan | |
| 1.1 | Pendahuluan | 1 |
| 1.2 | Latar Belakang Kajian | 2 |
| 1.3 | Penyataan Masalah | 6 |
| 1.4 | Persoalan Kajian | 8 |
| 1.5 | Objektif Kajian | 8 |
| 1.6 | Skop Kajian | 8 |
| 1.7 | Kepentingan Kajian | 10 |
| 1.8 | Organisasi Tesis | 11 |
| | | |
| BAB II | Kajian Literasi | |
| 2.1 | Pengenalan | 12 |
| 2.2 | Prestasi Dalam sektor Pendidikan Tinggi | 12 |
| | 2.2.1 Pengukuran Prestasi | 13 |
| | 2.2.2 Faktor Pengaruh Prestasi | 14 |
| | 2.2.3 Kaedah Analisis Prestasi | 17 |
| 2.3 | Pemahaman Penilaian Program Akreditasi | 18 |
| 2.4 | Perlombongan Data | 19 |
| | 2.4.1 Pemilihan Fitur | 19 |
| | 2.4.2 Teknik Pemilihan Fitur Dalam Kajian Ramalan Prestasi Pendidikan Tinggi | 27 |
| | 2.4.3 Pengelasan | 29 |

| | | |
|----------------------------------|---|----|
| 2.4.4 | Teknik Pengelasan Dalam Kajian Berkaitan Ramalan Prestasi Pendidikan Tinggi | 36 |
| 2.5 | Kesimpulan | 38 |
| BAB III METODOLOGI KAJIAN | | |
| 3.1 | Pengenalan | 39 |
| 3.2 | Perisian Yang Digunakan | 40 |
| 3.3 | Kerangka Metodologi Kajian | 40 |
| 3.4 | Fasa 1: Pemahaman bisnes | 41 |
| 3.4.1 | Mengenal pasti masalah, objektif dan skop | 41 |
| 3.4.2 | Kajian literasi | 43 |
| 3.5 | Fasa 2: Pemahaman Data | 43 |
| 3.5.1 | Pengumpulan data | 43 |
| 3.5.2 | Penerokaan data | 45 |
| 3.6 | Fasa 3: Penyediaan Data | 48 |
| 3.6.1 | Pembersihan data | 48 |
| 3.6.2 | Analisis deskriptif | 53 |
| 3.6.3 | Mengendali data tidak seimbang. | 66 |
| 3.7 | Fasa 4: Pemodelan | 67 |
| 3.7.1 | Pemilihan fitur | 68 |
| 3.7.2 | Pemilihan Pengelasan | 71 |
| 3.8 | Fasa 5: Penilaian | 72 |
| 3.9 | Fasa 6: Pelaksanaan | 74 |
| 3.10 | Kesimpulan | 74 |
| BAB IV PEMILIHAN FITUR | | |
| 4.1 | Pengenalan | 75 |
| 4.2 | Teknik Pemilihan Fitur | 75 |
| 4.2.1 | Teknik Penapis <i>Information Gain</i> | 75 |
| 4.2.2 | Teknik pembalut Pemilihan Berjjukan ke Depan | 78 |
| 4.2.3 | Teknik pembalut Penyingkiran Berjjukan ke Belakang | 79 |
| 4.2.4 | Teknik Carian Rawak Algoritma Genetik | 80 |
| 4.2.5 | Teknik Hibrid Penapis-Pembalut PBD. | 81 |
| 4.3 | Perbandingan Subset Fitur Terpilih | 82 |
| 4.4 | Kesimpulan | 83 |

| | | |
|-----------------|-----------------------------------|-----------|
| BAB V | MODEL PENGELASAN | |
| 5.1 | Pengenalan | 85 |
| 5.2 | Model Pengelasan Regresi logistik | 86 |
| 5.3 | Model Pengelasan Naive Bayes | 87 |
| 5.4 | Model Pengelasan Hutan Rawak | 89 |
| 5.5 | Pemilihan Model Pengelasan | 90 |
| 5.6 | Perbincangan Model Terpilih | 92 |
| 5.7 | Kesimpulan | 94 |
| | | |
| BAB VI | RUMUSAN DAN CADANGAN | |
| 6.1 | Pengenalan | 95 |
| 6.2 | Rumusan Kajian | 95 |
| 6.3 | Sumbangan Kajian | 96 |
| 6.4 | Batasan Kajian | 97 |
| 6.5 | Cadangan Kajian Masa Hadapan | 97 |
| | | |
| RUJUKAN | | 98 |
| | | |
| LAMPIRAN | | |
| Lampiran A | Senarai Fitur Asal | 106 |

SENARAI JADUAL

| No. Jadual | | Halaman |
|-------------------|--|----------------|
| Jadual 1.1 | Statistik penerimaan program PA dan FA berdasarkan tahun penerimaan dokumen. | 5 |
| Jadual 2.1 | Teknik pemilihan fitur untuk ramalan prestasi pendidikan tinggi | 28 |
| Jadual 2.2 | Teknik pengelasan untuk ramalan prestasi pendidikan tinggi | 37 |
| Jadual 3.1 | Senarai perisian yang digunakan | 40 |
| Jadual 3.2 | Jadual laporan kualiti data | 43 |
| Jadual 3.3 | Senarai fitur tidak berkepentingan | 45 |
| Jadual 3.4 | Senarai bidang diinput secara manual | 51 |
| Jadual 3.5 | Senarai fitur baharu | 52 |
| Jadual 3.6 | Penjanaan fitur kelas sasaran | 52 |
| Jadual 3.7 | Senarai fitur setelah aktiviti pembersihan data | 53 |
| Jadual 3.8 | Senarai analisis deskriptif | 54 |
| Jadual 3.9 | Formula metrik kekeliruan | 73 |
| Jadual 4.1 | Keputusan skor ujian IG | 76 |
| Jadual 4.2 | Tetapan parameter PK | 77 |
| Jadual 4.3 | Senarai fitur terpilih kaedah penapis | 78 |
| Jadual 4.4 | Senarai fitur terpilih kaedah pembalut teknik PBD | 79 |
| Jadual 4.5 | Senarai fitur terpilih kaedah pembalut teknik SBB | 79 |
| Jadual 4.6 | Tetapan parameter carian rawak AG | 80 |
| Jadual 4.7 | Senarai fitur terpilih teknik carian rawak AG | 80 |
| Jadual 4.8 | Senarai fitur terpilih teknik hibrid penapis-pembalut PBD | 81 |
| Jadual 4.9 | Perbandingan fitur terpilih berdasarkan teknik pemilihan fitur | 82 |
| Jadual 4.10 | Perbandingan prestasi teknik pemilihan fitur | 83 |
| Jadual 5.1 | Langkah pengelas RL berfungsi | 86 |

| | | |
|-------------|---------------------------------------|----|
| Jadual 5.2 | Tetapan parameter pengelas RL | 86 |
| Jadual 5.3 | Prestasi model pengelas RL | 87 |
| Jadual 5.4 | Langkah pengelas NB berfungsi | 87 |
| Jadual 5.5 | Prestasi model pengelas NB | 88 |
| Jadual 5.6 | Langkah pengelas HR berfungsi | 89 |
| Jadual 5.7 | Tetapan parameter pengelas HR | 89 |
| Jadual 5.8 | Prestasi model pengelas HR | 90 |
| Jadual 5.9 | Perbandingan prestasi model pengelas | 90 |
| Jadual 5.10 | Kepentingan subset fitur yang dipilih | 92 |
| Jadual 6.1 | Subset fitur optimum yang terpilih | 96 |

Pusat Sumber
FTSM

SENARAI ILUSTRASI

| No. Rajah | | Halaman |
|------------------|---|----------------|
| Rajah 1.1 | Matlamat Teras MQA | 2 |
| Rajah 1.2 | Jenis Tahap dan Kelayakan MQF | 4 |
| Rajah 2.1 | Proses menghasilkan subset fitur strategi carian menyeluruh | 23 |
| Rajah 2.2 | Proses pemilihan fitur strategi PBD | 24 |
| Rajah 2.3 | Proses pemilihan fitur strategi SBB | 25 |
| Rajah 2.4 | Ilustrasi lokal optima dan global optima | 26 |
| Rajah 2.5 | Contoh Model Pengelasan PK | 30 |
| Rajah 2.6 | Formula <i>Euclidean</i> | 31 |
| Rajah 2.7 | Contoh Model Pengelasan k-JT | 31 |
| Rajah 2.8 | Ilustrasi MSV | 33 |
| Rajah 2.9 | Ilustrasi RL | 34 |
| Rajah 2.10 | Ilustrasi model pengelas HR | 34 |
| Rajah 2.11 | Ilustrasi Rangkaian Saraf Buatan | 35 |
| Rajah 3.1 | Kaedah Perlombongan Data CRISP-DM | 39 |
| Rajah 3.2 | Kerangka Metodologi Kajian | 41 |
| Rajah 3.3 | Proses penilaian program Akreditasi Provisional (PA) | 42 |
| Rajah 3.4 | Proses penilaian program Akreditasi Penuh (FA) | 42 |
| Rajah 3.5 | Taburan permohonan program | 46 |
| Rajah 3.6 | Peratus permohonan akreditasi PA dan FA | 46 |
| Rajah 3.7 | Jumlah permohonan berdasarkan kategori PPT | 47 |
| Rajah 3.8 | Peratus permohonan berdasarkan negeri PPT | 47 |
| Rajah 3.9 | Kod pengaturcaraan SQL membuang data tidak lengkap dan tidak betul. | 49 |
| Rajah 3.10 | Nilai yang hilang bagi fitur 'Kod merah'. | 50 |

| | | |
|------------|--|----|
| Rajah 3.11 | Kod pengaturcaraan SQL untuk mengisi nilai data fitur 'Kod merah' yang hilang. | 50 |
| Rajah 3.12 | Kod pengaturcaraan SQL untuk mengisi nilai data fitur 'Penilaian' yang hilang. | 50 |
| Rajah 3.13 | Kod pengaturcaraan SQL untuk penyeragaman fitur 'Tahap'. | 51 |
| Rajah 3.14 | Peratus keseluruhan prestasi | 55 |
| Rajah 3.15 | Statistik prestasi mengikut jenis akreditasi | 55 |
| Rajah 3.16 | Statistik prestasi mengikut jenis permohonan | 56 |
| Rajah 3.17 | Statistik prestasi mengikut tahap & peringkat | 57 |
| Rajah 3.18 | Statistik prestasi mengikut bidang NEC | 58 |
| Rajah 3.19 | Statistik penerimaan dokumen mengikut bulan | 59 |
| Rajah 3.20 | Statistik prestasi mengikut kategori kod merah | 60 |
| Rajah 3.21 | Statistik prestasi mengikut status | 61 |
| Rajah 3.22 | Tempoh (minggu) pendaftaran oleh UP | 62 |
| Rajah 3.23 | Tempoh (minggu) semakan oleh US | 62 |
| Rajah 3.24 | Peratus prestasi mengikut bahagian | 63 |
| Rajah 3.25 | Tempoh (minggu) penilaian oleh panel penilai | 64 |
| Rajah 3.26 | Statistik penerimaan mengikut panel penilai | 64 |
| Rajah 3.27 | Statistik prestasi mengikut kategori PPT | 65 |
| Rajah 3.28 | Statistik prestasi mengikut taraf PPT | 66 |
| Rajah 3.29 | Teknik pensampelan SMOTE | 67 |
| Rajah 3.30 | Proses pemilihan fitur kaedah penapis | 68 |
| Rajah 3.31 | Proses pemilihan fitur kaedah pembalut | 69 |
| Rajah 3.32 | Proses pemilihan fitur strategi carian Algoritma Genetik (AG) | 70 |
| Rajah 3.33 | Proses pemilihan fitur kaedah hibrid | 71 |
| Rajah 3.34 | Metrik kekeliruan | 72 |
| Rajah 4.1 | Keputusan eksperimen kaedah penapis | 78 |

Pusat Sumber
FTSM

SENARAI SINGKATAN

| | |
|----------|--------------------------------------|
| AG | Algoritma Genetik |
| BDA | Berjukkan Dua Arah |
| CAR | Nisbah Kecukupan Modal |
| CIR | Nisbah Kos Pendapatan |
| DEA | Analisis Penyampulan Data |
| FN | <i>False Negative</i> |
| FP | <i>False Positive</i> |
| FTSM | Fakulti Teknologi dan Sains Maklumat |
| GI | Gini Indeks |
| HR | Hutan Rawak |
| IG | <i>Information Gain</i> |
| IPTA | Institusi Pendidikan Tinggi Awam |
| IPTS | Institusi Pendidikan Tinggi Swasta |
| k-JT | k-Jiran Terdekat |
| LTD | Hutang Jangka Panjang |
| MJA | Mesyuarat Jawatankuasa Akreditasi |
| MPK | Mesin Penggalak Kecerunan |
| MQA | Agensi Kelayakan Malaysia |
| MQF | Kerangka Kelayakan Malaysia |
| MSV | Mesin Sokongan Vektor |
| NB | <i>Naive Bayes</i> |
| NPL | Pinjaman Tidak Berbayar |
| NEC-2020 | Kod Pendidikan Nasional 2020 |
| PBD | Pemilihan Berjukkan ke Depan |
| PCR | Penyingkiran Ciri Rekursif |

| | |
|-------|---|
| PK | Pokok Keputusan |
| PPT | Pemberi Pendidikan Tinggi |
| RL | Regresi Logistik |
| RMSE | <i>Root Mean Squared Error</i> |
| ROC | <i>Receiver Operating Characteristic</i> |
| ROA | Pulangan Ke Atas Aset |
| ROE | Pulangan Ke Atas Ekuiti |
| RSB | Rangkaian Saraf Buatan |
| SBB | Penyingkiran Berjjukan ke Belakang |
| SMOTE | <i>Synthetic Minority Over-sampling Technique</i> |
| TP | <i>True Positive</i> |
| TN | <i>True Negative</i> |
| UKM | Universiti Kebangsaan Malaysia |
| UP | Unit Pendaftaran |
| US | Unit Semakan |

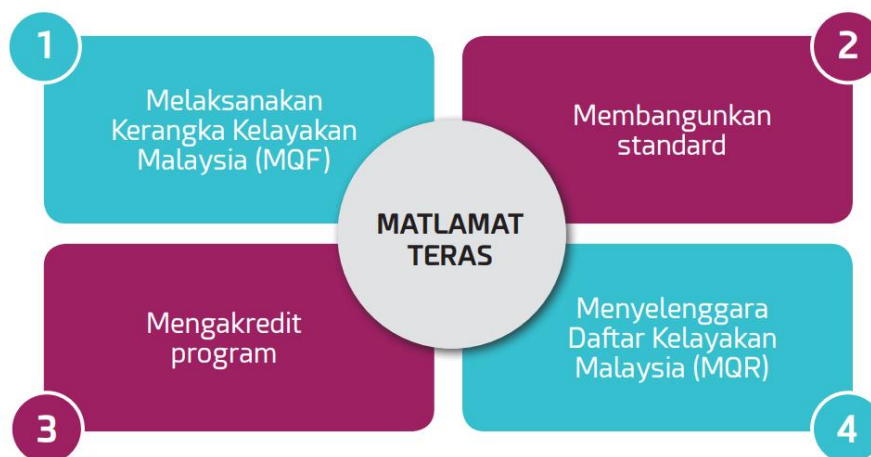
BAB I

PENGENALAN

1.1 PENDAHULUAN

Amalan terbaik jaminan kualiti pendidikan tinggi memerlukan kriteria dan standard yang jelas, telus dan adil untuk menjadi rujukan bagi penilaian dan laporan akreditasi program yang ditawarkan oleh Penyedia Pendidikan Tinggi (PPT). Akreditasi program ialah jaminan ke atas kualiti program yang ditawarkan oleh PPT bagi memastikan program yang ditawarkan mencapai kualiti yang diperlukan dengan mengikut kod amalan dan standard yang ditetapkan. Terdapat dua proses akreditasi program iaitu Akreditasi Provisional (PA) yang menentukan sama ada sesuatu program telah memenuhi keperluan kualiti minimum yang menjadi syarat kepada Akreditasi Penuh dan Akreditasi Penuh (FA) iaitu penilaian untuk memastikan bahawa pengajaran, pembelajaran dan semua aktiviti berkaitan program yang disediakan oleh PPT telah memenuhi kualiti dan mematuhi Kerangka Kelayakan Malaysia (MQF) (Portal Rasmi MQA).

PPT yang berjaya menawarkan program yang berkualiti dilihat dapat menyumbang kepada pembangunan modal insan, masyarakat berpengetahuan, daya saing negara dan menjadi pusat kecemerlangan pengajian tinggi. Bagi memastikan PPT dapat menawarkan program pengajian, program perlulah diakredit seperti yang dinyatakan pada matlamat teras ketiga di dalam Pelan Strategik MQA 2021-2025. Program yang melalui penilaian perlulah mematuhi standard, kod amalan dan memastikan proses jaminan kualiti memenuhi tempoh piagam yang ditetapkan agar program berkualiti dapat ditawarkan (Perancangan Strategik MQA 2021-2025).



Rajah 1.1 Matlamat Teras MQA

Sumber: Perancangan Strategik MQA 2021-2025

Tempoh piagam pelanggan bagi proses akreditasi program adalah bermula dari tarikh dokumen lengkap hingga tarikh Mesyuarat Jawatankuasa Akreditasi (MJA). Tempoh piagam ini berbeza bagi penilaian Akreditasi Provisional (PA) dan Akreditasi Penuh (FA) iaitu 3 bulan 3 minggu bagi PA dan 7 bulan bagi FA.

Tempoh piagam ini telah digunakan bermula tahun 2011. Pada tahun 2018, kerajaan telah menyarankan agar semua kementerian dan agensi kerajaan melihat semula proses dan tempoh piagam pelanggan bagi menambahbaik mutu perkhidmatan yang diberikan. Cadangan tersebut disambut baik dan tempoh piagam penilaian program PA dan FA telah dikaji bagi meningkatkan kepuasan pelanggan. Bermula Julai 2019 tempoh piagam sedia ada telah dipendekkan kepada sembilan (9) minggu bagi tempoh dari tarikh dokumen lengkap hingga tarikh Mesyuarat Jawatankuasa Akreditasi (MJA). Namun, jumlah minggu keseluruhan proses penilaian termasuk minggu yang tidak diambil kira dalam kiraan piagam pelanggan adalah selama 14 minggu (Lampiran A: Pekeliling MQA Bil. 6.2019).

1.2 LATAR BELAKANG KAJIAN

Kajian ini dikategorikan dalam bidang pengecaman pola dan sokongan keputusan. Kajian ini menggunakan pendekatan analitik ramalan untuk mengenal pasti fitur yang penting yang dapat mengecam pola dalam permohonan akreditasi yang bermasalah seperti status permohonan ditolak dan tempoh penilaian yang diambil oleh panel penilai

yang membuat penilaian program berdasarkan data sejarah dan arah aliran. Dengan mengenali pola ini, pihak pengurusan atau pembuat keputusan boleh melaksanakan intervensi untuk menangani isu yang berpotensi timbul semasa penilaian program dan meningkatkan peluang kejayaan untuk permohonan akreditasi yang bermasalah. Analitik ramalan boleh memberikan sokongan keputusan dengan menawarkan cerapan tentang kemungkinan kejayaan atau kegagalan suatu permohonan berdasarkan data sejarah. Maklumat ini boleh membantu pembuat keputusan mengutamakan sumber, memperuntukkan pembiayaan dan membuat keputusan termaklum untuk campur tangan terhadap permohonan program akreditasi yang bermasalah.

Menyedari kepentingan akreditasi, proses penilaian yang dilaksanakan ke atas program yang dipohon oleh PPT perlulah dapat diselesaikan dalam tempoh yang ditetapkan iaitu 14 minggu. Faktor-faktor yang mempengaruhi tempoh proses penilaian akan dikaji merangkumi analisis ke atas penerimaan permohonan penilaian PA dan FA. Tempoh penilaian boleh dipengaruhi oleh pelbagai faktor atau fitur yang boleh memberi kesan kepada masa yang diambil untuk menilai permohonan. Terdapat faktor-faktor yang diketahui dan terdapat juga faktor yang tidak diketahui. Antara faktor yang diperhatikan termasuk:

1. Kelengkapan Maklumat: Kesempurnaan dan ketepatan maklumat yang diberikan oleh PPT boleh memberi kesan yang ketara kepada masa penilaian. Maklumat yang tidak lengkap atau tidak tepat mungkin memerlukan maklum balas atau pengesahan tambahan daripada pemohon seterusnya memanjangkan tempoh proses penilaian.
2. Kerumitan: Kerumitan termasuk jenis permohonan atau jenis program sama ada akreditasi provisional atau akreditasi penuh, dan tahap perincian yang diperlukan boleh mempengaruhi masa penilaian. Permohonan penilaian program daripada PPT juga adalah dari pelbagai tahap, kelayakan dan bidang. Terdapat lapan (8) tahap dari pelbagai kelayakan yang digariskan di dalam MQF Edisi Kedua (Agensi Kelayakan Malaysia 2018) seperti Rajah 1.2 di bawah.

| MQF Level | Minimum Graduating Credit | Academic Sector | TVET Sector |
|-----------|---------------------------|--|----------------------|
| 8 | No credit rating | PhD by Research | |
| | 80 | Doctoral Degree by Mixed Mode & Coursework | |
| 7 | No credit rating | Master's Degree by Research | |
| | 40 | Master's Degree by Mixed Mode & Coursework | |
| | 30 | Postgraduate Diploma | |
| | 20 | Postgraduate Certificate | |
| 6 | 120 | Bachelor's degree | Bachelor's degree |
| | 64* | Graduate Diploma | Graduate Diploma |
| | 34* | Graduate Certificate | Graduate Certificate |
| 5 | 40 | Advanced Diploma | Advanced Diploma |
| 4 | 90 | Diploma | Diploma |
| 3 | 60 | Certificate | Certificate |
| 2 | 30 | Certificate | Certificate |
| 1 | 15 | Certificate | Certificate |

* Inclusive of 4 credits from general studies subjects.

Rajah 1.2 Jenis Tahap dan Kelayakan MQF

Sumber: Kerangka Kelayakan Malaysia Edisi Kedua

- Jumlah Permohonan:** Jumlah bilangan permohonan yang perlu diproses secara serentak boleh menjejaskan masa penilaian. Jumlah permohonan yang tinggi boleh mengakibatkan masa pemprosesan yang lebih lama disebabkan oleh kekangan sumber dan pertimbangan keutamaan. Setiap tahun, terdapat banyak permohonan yang diterima daripada PPT bagi penilaian program akreditasi dan jumlahnya adalah tidak menentu. Ada masanya berlaku peningkatan permohonan program pada satu-satu masa. Menurut statistik penerimaan program baharu seperti di Jadual 1.1, peningkatan sebanyak 23% permohonan program akreditasi diterima pada tahun 2021 iaitu 1247 program berbanding 1014 program yang diterima pada tahun 2020. Jumlah permohonan ini memberi kesan langsung kepada agihan tugas pegawai untuk menyelesaikan proses penilaian program dalam tempoh yang ditetapkan.

Jadual 1.1 Statistik penerimaan program PA dan FA berdasarkan tahun penerimaan dokumen.

| Tahun | PA | FA | Jumlah | Catatan |
|-------|-----|-----|--------|--------------------|
| 2019 | 228 | 672 | 900 | Bermula Julai 2019 |
| 2020 | 423 | 591 | 1014 | |
| 2021 | 476 | 771 | 1247 | |
| 2022 | 440 | 547 | 987 | |

4. **Keperluan Kawal Selia:** Pematuhan kepada keperluan kawal selia dan dasar dalaman boleh menambah masa kepada proses penilaian. Memastikan semua piawaian peraturan dipenuhi mungkin akan melibatkan langkah dokumentasi dan pengesahan tambahan.
5. **Kualiti Dokumentasi:** Kualiti dan organisasi dokumentasi sokongan yang dikemukakan bersama permohonan boleh memberi kesan kepada masa penilaian. Permohonan yang diterima bagi penilaian akreditasi program adalah daripada semua negeri di Malaysia dengan pelbagai kategori PPT yang memohon seperti Universiti Awam (UA), Institusi Pengajian Tinggi Swasta (IPTS), Politeknik dan taraf PPT yang berbeza seperti universiti, politeknik, kolej dan pusat latihan. Perbezaan PPT ini menyebabkan dokumentasi yang disediakan berbeza-beza. Dokumentasi yang disediakan dengan baik dan mudah diakses boleh mempercepatkan proses semakan dan penilaian.
6. **Kepelbagaian Bidang:** Terdapat 10 bidang NEC dan 140 kategori bidang pengkhususan NEC di dalam Kod Pendidikan Nasional 2020 yang disenaraikan sebagai kod rujukan standard bidang pendidikan dan latihan di negara ini sehingga sekarang (Kementerian Pengajian Tinggi Malaysia 2021). Kepelbagaian tahap, kelayakan dan kod bidang program ini memerlukan penilaian dilaksanakan oleh panel penilai. Ada masanya terdapat kekangan untuk mendapatkan kepakaran panel penilai khususnya bagi permohonan program bidang baharu. Bagi bidang-bidang tertentu pula, penilaian dilaksanakan oleh panel penilai badan kualiti lain seperti badan profesional yang mempunyai autoriti berdasarkan ketetapan akta masing-masing.

Mengetahui pola permohonan akreditasi yang bermasalah atau tidak, membolehkan organisasi melakukan intervensi atau berusaha ke arah mengurangkan masa yang diambil untuk menilai permohonan akreditasi dan meningkatkan kecekapan keseluruhan organisasi. Persoalannya adakah benar faktor ini boleh digunakan dalam meramal dan mengoptimumkan proses penilaian?

Oleh itu kajian ini dilaksanakan bagi memeriksa faktor-faktor yang mempengaruhi prestasi penilaian program akreditasi baharu sama ada dapat dilaksanakan di dalam tempoh atau tidak dan seterusnya meramal prestasi pada masa hadapan. Terdapat beberapa pendekatan yang sistematik dan komprehensif yang boleh digunakan untuk mengesahkan faktor yang mempengaruhi tempoh penilaian permohonan akreditasi seperti temu bual, pemetaan proses dan analisis data. Kajian ini menggunakan pendekatan pemilihan fitur untuk mengenal pasti faktor paling relevan yang mempengaruhi tempoh penilaian program akreditasi. Pemilihan fitur ialah proses di mana subset fitur yang berkaitan dipilih daripada set fitur asal berdasarkan kepentingannya dalam meramalkan pemboleh ubah sasaran (dalam kajian ini, prestasi tempoh penilaian).

Dapatan daripada kajian ini dijangka dapat memberi pengetahuan baharu kepada komuniti bidang pengurusan dan sains data, khususnya dalam pemodelan untuk menambah baik proses kerja dan meningkatkan pembuatan keputusan.

1.3 PENYATAAN MASALAH

Peningkatan prestasi adalah amat penting kepada semua agensi atau jabatan. Kajian berkenaan prestasi bagi institusi pendidikan tinggi telah banyak dijalankan antaranya kajian berkenaan prestasi graduan (Gunawan & Palupiningsih 2017) (Yohana et al. 2021) dan akreditasi universiti (Musthafa et al. 2020; Noripansyah et al. 2023). Faktor-faktor yang mempengaruhi prestasi tempoh penilaian dikaji bagi langkah-langkah intervensi awal atau penambahbaikan.

Tugas mengenal pasti faktor-faktor yang menyumbang kepada permohonan yang bermasalah atau tertangguh menimbulkan cabaran yang ketara kerana faktor-faktor ini tidak diketahui dan tidak pernah dikaji secara sistematik menggunakan

pendekatan sains data. Sehingga disertasi ini ditulis, belum ada kajian khusus di Malaysia yang mengkaji faktor atau pola yang boleh menentukan sama ada sesuatu program pengajian akreditasi baharu yang dipohon oleh PPT dapat diproses di dalam tempoh atau di luar tempoh yang ditetapkan.

Dalam konteks ini, kaedah tradisional tidak mencukupi dalam mendedahkan faktor kritikal dengan cekap. Oleh itu, terdapat keperluan mendesak untuk menggunakan pembelajaran mesin dan teknik pengoptimuman (kaedah pemilihan fitur) untuk mengekstrak subset fitur terbaik daripada set data yang boleh diperolehi dari organisasi terbabit. Dengan menyepadukan metodologi lanjutan ini, kajian ini menyasarkan untuk membangunkan model ramalan yang boleh mengenal pasti secara berkesan faktor utama yang berkaitan dengan permohonan yang bermasalah atau tertangguh, dengan itu meningkatkan ketepatan dan kecekapan sistem pemrosesan permohonan program.

Berdasarkan sumber data yang ada, fitur yang berkaitan perlu diuji sama ada fitur ini mempengaruhi prestasi tempoh penilaian program akreditasi baharu. Kekurangan sumber data dan sumber data yang tidak seimbang boleh menyebabkan bias terhadap kelas majoriti menyebabkan model yang dibangunkan perlu kukuh (*robust*) untuk mengelaskan tempoh penilaian dengan prestasi yang tinggi.

Selaras dengan Teorem Tiada Makan Tengah Hari Percuma (*No Free Lunch Theorem*) (Wolpert & Macready 1997) yang menyatakan bahawa tiada satu algoritma berprestasi terbaik untuk semua masalah. Dalam erti kata lain, tidak ada satu pun model pembelajaran mesin yang sesuai digunakan untuk setiap situasi kajian atau setiap jenis data set dalam setiap domain kajian. Model yang terbaik untuk satu kajian mungkin tidak sesuai untuk kajian yang lain. Ia menunjukkan bahawa tidak ada algoritma pembelajaran yang sempurna (Sterkenburg & Grünwald 2021). Oleh itu, terdapat keperluan untuk menguji beberapa algoritma pembelajaran mesin, teknik pengoptimuman dan kaedah pemilihan fitur untuk mengekstrak subset fitur yang paling bermaklumat daripada data yang tersedia. Dengan mengguna pakai pendekatan pelbagai rupa berpandukan prinsip teorem tersebut, kajian ini berhasrat untuk membangunkan model ramalan yang utuh yang mampu mengenal pasti faktor utama

yang berkaitan dengan permohonan yang bermasalah atau tertangguh, sekali gus meningkatkan kecekapan dan ketepatan sistem pemrosesan permohonan akreditasi.

1.4 PERSOALAN KAJIAN

Permasalahan dalam penilaian program akreditasi baharu adalah perkara yang menarik minat untuk dikaji. Beberapa persoalan berkenaan kajian ini telah dikenal pasti iaitu:

1. Apakah fitur penting yang boleh digunakan untuk membina model ramalan prestasi penilaian program akreditasi baharu?
2. Apakah algoritma pembelajaran mesin yang sesuai untuk menghasilkan nilai prestasi terbaik bagi model pengelasan prestasi penilaian program akreditasi baharu?

1.5 OBJEKTIF KAJIAN

Kajian ini secara khususnya ingin mencapai objektif-objektif seperti berikut:

1. Untuk mengenal pasti fitur penting dalam meramal prestasi program akreditasi baharu;
2. Merekayasa dan membangunkan model pengelasan prestasi penilaian program akreditasi.

1.6 SKOP KAJIAN

Kajian ini menggunakan set data yang diperolehi secara kelulusan dari agensi yang mengendalikan proses penilaian program akreditasi di Malaysia. Bagi tujuan kajian ini, prestasi penilaian program akreditasi baharu ditakrifkan sebagai pelaksanaan proses penilaian program akreditasi yang baharu diterima daripada PPT sama ada dapat dilaksanakan di dalam tempoh atau di luar tempoh yang ditetapkan. Pelaksanaan kajian bagi pemilihan fitur dan pengelasan kajian ini akan menggunakan perisian *RapidMiner*. Setiap hasil daripada analisis dan keputusan kajian akan dinilai melalui proses penilaian untuk memilih fitur dengan model pengelasan yang terbaik.

Set data ini mempunyai sebanyak 5024 rekod data dengan 28 fitur yang merupakan set data penerimaan program baharu bagi tujuan penilaian Akreditasi Provisional (PA) dan Akreditasi Penuh (FA) yang diterima dari 1 Julai 2019 sehingga Disember 2022. Hanya program yang telah selesai dinilai sahaja dipilih untuk tujuan analisis kerana semua peringkat proses telah termasuk di dalam penilaian ini bermula dari penerimaan permohonan program sehingga mesyuarat keputusan penilaian program.

Sebanyak lima teknik pemilihan fitur digunakan bagi memilih subset fitur yang optimum iaitu (1) teknik penapis *Information Gain* IG, (2) teknik pambalut Pemilihan Berjjukan ke Depan PBD, (3) teknik pambalut Penyingkiran Berjjukan ke Belakang SBB, (4) teknik carian rawak Algoritma Genetik AG dan (5) teknik hibrid penapis-pambalut PBD. Teknik penapis dan teknik pambalut PBD mempunyai kelebihan dari segi kecekapan. Teknik ini banyak digunakan di dalam kajian untuk mencari fitur yang mempengaruhi prestasi. Teknik pambalut SBB pula dapat mengurangkan terlebih padanan dan membuang fitur yang berulang manakala teknik carian rawak AG dapat memilih fitur yang penting dan subset fitur optimum yang boleh meningkatkan prestasi model pembelajaran mesin. Teknik hibrid dipilih bagi memanfaatkan kelebihan teknik pemilihan fitur yang berbeza.

Tiga teknik algoritma pengelasan pembelajaran mesin dipilih bagi meramal prestasi iaitu (1) Regresi Logistik RL, (2) *Naive Bayes* NB dan (3) Hutan Rawak HR. Ketiga-tiga model ini akan diuji menggunakan fitur-fitur yang telah dipilih melalui eksperimen pemilihan fitur. RL merupakan algoritma yang sesuai dengan pengelasan binari. RL mempunyai kelebihan dari segi kepantasan untuk set data latihan walaupun bagi set data yang besar. NB berfungsi pada konsep yang mudah dan sederhana tetapi relevan yang mana dalam beberapa kes, *Bayes* mengatasi prestasi banyak algoritma yang agak kompleks. HR pula dipilih kerana mengurangkan terlebih padanan dan boleh menghasilkan keputusan ketepatan ramalan yang lebih baik. Ketiga-tiga algoritma pengelasan ini juga banyak digunakan di dalam kajian peramalan prestasi seperti yang dinyatakan di dalam kajian literasi di dalam Bab II.

Prestasi setiap model yang dibangunkan, dinilai dan dibandingkan berdasarkan kompleksiti model yang diukur berdasarkan pemilihan fitur dan prestasi model yang merangkumi ketepatan (*accuracy*), dapatan semula (*recall*), kejitian (*precision*) dan skor-f. Skor-f merupakan pengukuran harmoni diantara dapatan semula dan kejitian. Model pengelasan yang mempunyai skor-f dengan kompleksiti fitur terbaik dipilih sebagai model pengelasan paling berpotensi untuk kajian ramalan prestasi penilaian program akreditasi baharu.

1.7 KEPENTINGAN KAJIAN

Penyelidikan yang digariskan dalam pernyataan masalah mempunyai kepentingan yang signifikan kepada komuniti sains data dan bidang pengurusan. Hasil kajian ini dijangka dapat memberi sumbangan dan impak berikut:

1. Komuniti bidang pengurusan: Komuniti atau organisasi pengurusan, khususnya yang mengawal kualiti program pendidikan tinggi dapat mengenal pasti fitur yang dominan dalam mempengaruhi prestasi penilaian akreditasi. Oleh itu pihak pengurusan dapat menyusun strategi bagi meningkatkan kecekapan seterusnya prestasi dan produktiviti agensi. Kajian ini amat penting untuk dilaksanakan kerana buat masa ini belum ada kajian yang seumpama dengannya dijalankan;
2. Pihak berkepentingan: Kajian ini menghasilkan pengetahuan baharu agar organisasi dapat mengkaji dan menambah baik proses penilaian program akreditasi;
3. Dengan menyepadukan kaedah pemilihan fitur, teknik pengoptimuman dan pembelajaran mesin untuk menangani masalah dunia sebenar dengan faktor yang tidak diketahui, penyelidikan ini menyumbang kepada kemajuan metodologi dalam bidang sains data. Kajian ini mempamerkan potensi menggabungkan pendekatan yang pelbagai untuk menangani cabaran kompleks dengan berkesan.
4. Kajian ini menangani cabaran dunia sebenar yang praktikal dan penting yang dihadapi dalam pelbagai sistem pemrosesan permohonan. Penemuan dan

pandangan yang diperoleh daripada penyelidikan ini mempunyai implikasi praktikal untuk industri dan organisasi yang bergantung pada pemrosesan permohonan yang cekap.

5. Kajian ini membangunkan model ramalan untuk mengenal pasti faktor-faktor yang membawa kepada permohonan bermasalah atau tertunda yang boleh meningkatkan ketepatan dan kecekapan pemodelan ramalan dalam pelbagai aplikasi. Ini boleh membawa kepada proses membuat keputusan yang lebih baik dan prestasi operasi yang lebih efisien.

1.8 ORGANISASI TESIS

Tesis ini terbahagi kepada lima bab. Bab I menerangkan kajian yang dilaksanakan iaitu berkaitan latar belakang kajian, pernyataan masalah, objektif kajian, kepentingan kajian dan skop kajian secara keseluruhan. Fokus Bab II ialah kajian-kajian terdahulu berkenaan konsep dan teknik yang berkaitan dengan akreditasi dan prestasi, pemilihan fitur dan teknik pembelajaran mesin dalam meramal prestasi model pengelasan. Di dalam Bab III pula, metodologi kajian yang dilaksanakan di dalam kajian ini akan dihuraikan dengan terperinci. Ia merangkumi fasa memahami data, pra pemrosesan data dan mengenal pasti ciri-ciri dan struktur data. Selain itu, reka bentuk kajian juga diterangkan dengan lebih lanjut di dalam bab ini. Bab IV dan Bab V memperincikan analisis hasil pemilihan fitur dan pembangunan model pembelajaran mesin serta perbandingan prestasi bagi keputusan eksperimen yang dijalankan. Bab akhir merupakan rumusan, sumbangan dan cadangan penambahbaikan yang boleh dibuat pada masa akan datang berdasarkan kajian yang telah dilaksanakan ini.

BAB II

KAJIAN LITERASI

2.1 PENGENALAN

Secara umumnya bab ini menerangkan kajian kesusasteraan yang pernah dilaksanakan berkenaan prestasi khususnya yang berkaitan dengan pendidikan tinggi meliputi pengukuran prestasi, faktor-faktor yang mempengaruhi prestasi dan kaedah analisis data yang digunakan. Selain itu, pemilihan fitur dan teknik pengelasan khususnya berkaitan prestasi juga akan dihuraikan. Melalui pembacaan kajian terdahulu, belum menemui satu kajian di Malaysia yang mempunyai permasalahan kajian berkaitan pemilihan fitur dan pembelajaran mesin berkaitan pengelasan prestasi penilaian program akreditasi baharu. Namun, kajian kesusasteraan ini akan dibahagikan kepada beberapa bahagian iaitu:

1. Prestasi Dalam Sektor Pendidikan Tinggi;
2. Pemahaman Penilaian Program Akreditasi;
3. Perlombongan Data;
4. Kesimpulan.

2.2 PRESTASI DALAM SEKTOR PENDIDIKAN TINGGI

Menurut kamus Dewan Edisi Keempat, prestasi bermaksud hasil yang telah diperoleh atau pencapaian. Prestasi boleh merujuk kepada sejauh mana sistem, proses atau individu mencapai matlamat atau tugas yang dimaksudkan. Ia sering dikaitkan dengan kecekapan dan keberkesanan (Paschoalotto et al. 2020).

2.2.1 Pengukuran Prestasi

Pengukuran prestasi ialah satu proses mengukur dan menilai keberkesanan dan kecekapan (Moullin 2007) sesebuah organisasi atau individu dalam mencapai objektif atau matlamat mereka. Proses peningkatan kualiti akan berjalan lancar dan berterusan dengan sokongan sistem pengukuran prestasi kualiti yang baik. Apabila memperkenalkan sesuatu penambahbaikan, organisasi perlu mengetahui keberkesanannya kepada tahap kualiti output. Oleh itu sistem pengukuran yang sesuai perlu dirangka mengikut jenis-jenis keluaran atau perkhidmatan yang diberikan (Pengurusan Kualiti Menyeluruh Bagi Perkhidmatan Awam 1992).

Pengukuran prestasi adalah berbeza mengikut sektor. Contohnya prestasi di dalam perbankan diukur berdasarkan nisbah keuntungan iaitu ROA dan ROE (Abduh 2014). Selain nisbah keuntungan, menurut Assous (2022), amalan pengurusan (CIR dan nisbah kecekapan), aset dan pinjaman (NPL dan LTD), nisbah kecukupan modal (CAR) dan kecairan (nisbah semasa dan nisbah cepat) merupakan ukuran yang digunakan untuk menilai prestasi institusi perbankan.

Dalam bidang perubatan, pengukuran prestasi penjagaan kesihatan dianalisis dan dinilai berdasarkan aktiviti berkaitan penjagaan kesihatan. Tujuannya adalah untuk mengenal pasti peluang untuk mengurangkan kos, meningkatkan kualiti penjagaan kesihatan dan meningkatkan kecekapan penyampaian perkhidmatan. Menurut Turukmane et al. (2023) dan Naveenkumar S et al. (2021) antara teknologi yang telah dibangunkan bagi meningkatkan prestasi penyampaian perkhidmatan ialah dengan teknologi ramalan kesihatan pintar menggunakan mesin pembelajaran. Teknologi ini berjaya mengesan penyakit dengan merujuk kepada simptom yang dimaklumkan oleh pesakit. Dengan teknologi ini, kos penjagaan kesihatan dapat dikurangkan melalui pengesanan awal penyakit seterusnya mencegah masalah kesihatan daripada menjadi lebih teruk.

Dalam bidang pertanian, iklim dan produktiviti adalah berkait untuk mengukur prestasi hasil tanaman. Isu seperti pemilihan dan penggiliran tanaman, keperluan air, keperluan baja dan penjagaan perlu bersesuaian dengan hasil yang ingin dicapai. Oleh kerana faktor iklim persekitaran yang berbeza dan berubah-ubah, terdapat keperluan

untuk mempunyai teknik yang cekap untuk memudahkan penanaman dan membantu petani dalam pengeluaran dan pengurusan pertanian (Hina & Hasan, 2022).

Pengukuran prestasi dalam pendidikan tinggi pula dilihat dari beberapa aspek seperti aspek pelajar, pensyarah dan institusi. Aspek pelajar diukur melalui pencapaian akademik (Wang et al. 2010; Kartini et al. 2017), pembangunan kemahiran (Mohamad Razi et al. 2022) serta kecekapan tingkah laku dan sosial (Zuluaga-Ortiz et al. 2023).

Prestasi pensyarah pula diukur dari segi kecekapan dan keberkesanan pengajaran (Almufarreh et al. 2023) dan output penyelidikan (Zuluaga-Ortiz et al. 2023). Manakala pengukuran prestasi institusi diukur melalui kualiti pendidikan (Almufarreh et al. 2023), hasil pendidikan, ranking (Kumar et al. 2020) dan akreditasi (Musthafa et al. 2020).

Akreditasi merupakan salah satu kriteria yang diukur bagi menentukan prestasi institusi. Akreditasi bagi program diperlukan oleh Pemberi Pendidikan Tinggi (PPT) bagi memastikan PPT dapat menawarkan program yang berkualiti. Oleh itu, adalah penting untuk meningkatkan kecekapan, memudahkan dan mempercepatkan perkhidmatan sekaligus dapat memenuhi kehendak pihak berkepentingan (Perancangan Strategik MQA 2021-2025) khususnya dalam menyokong usaha PPT mendapatkan akreditasi program.

Dalam konteks kajian ini, pengukuran prestasi merujuk kepada pencapaian pelaksanaan proses penilaian program akreditasi dinilai di dalam tempoh 14 minggu penilaian yang ditetapkan. Pencapaian prestasi ini dapat membuktikan kecekapan dan keberkesanan proses penilaian yang dilaksanakan oleh organisasi.

2.2.2 Faktor Pengaruh Prestasi

Prestasi boleh dipengaruhi oleh pelbagai faktor antaranya faktor individu yang merangkumi kemahiran dan kecekapan, motivasi (Casiraghi et al. 2022), dan kesihatan. Faktor persekitaran atau faktor luaran pula adalah seperti persekitaran kerja atau tempat belajar dan sumber yang ada (Kamara & Dadhabai 2022). Faktor teknikal juga

memainkan peranan yang penting dalam mempengaruhi prestasi seperti peralatan (Kamara & Dadhabai 2022), teknologi dan perisian. Apabila faktor-faktor ini diurus dengan cekap, organisasi dapat mencapai prestasi yang optimum.

Apabila melihat kepada konteks pendidikan tinggi contohnya di universiti, faktor motivasi diri, disiplin diri, sokongan rakan sebaya, dan pendekatan pengajaran pensyarah menjadi antara faktor pencapaian akademik pelajar (Sikhwari et al. 2019). Faktor seperti prestasi akademik, ciri peribadi, latar belakang sosioekonomi (Fahrudy & Uyun 2022; Patel 2021; Sadiq & Ahmed 2019) dan faktor institusi (Bedregal-Alpaca et al. 2020) pula dikenalpasti mempengaruhi prestasi graduan untuk menamatkan pengajian.

Terdapat juga kajian yang dijalankan untuk meramal prestasi pelajar di universiti sebelum mereka melanjutkan pelajaran ke universiti seperti yang dilaksanakan oleh Mengash (2020). Kajian dijalankan bagi mencari kriteria kemasukan ke universiti yang paling tepat untuk meramal prestasi akademik pelajar di masa hadapan. Faktor gred purata akademik, pencapaian skor ujian kemasukan dan skor ujian bakat menjadi ukuran meramal prestasi pelajar di masa hadapan.

Selain daripada prestasi graduan, Musthafa et al. (2020) membuat kajian berkenaan dengan faktor-faktor yang mempengaruhi prestasi institusi untuk mengekalkan akreditasi iaitu data pelajar, akademik, kemasukan, dan alumni. Faktor prestasi institusi juga dipengaruhi oleh faktor kepemimpinan, budaya organisasi, dasar dan prosedur (SintayehuShibru 2017) manakala faktor berkaitan tugas sama ada dari segi kerumitan tugas dan kejelasan matlamat juga akan mempengaruhi prestasi.

Peruntukan sumber yang ditetapkan dalam sistem universiti dan kesannya terhadap prestasi di institusi pengajian tinggi telah dikaji oleh Liefner (2003). Dalam kajian beliau mendapati peruntukan berdasarkan prestasi, cenderung membawa kepada perubahan positif namun ia juga memberi kesan sampingan yang tidak diinginkan. Perubahan dalam peruntukan sumber memberi kesan kepada tahap dan jenis aktiviti yang boleh dilaksanakan di universiti.

Kajian yang dibuat oleh Leiber et al. (2018) pula mendapati akreditasi program tidak menunjukkan kesan yang ketara ke atas peningkatan kualiti program. Beberapa cabaran dikenal pasti dalam pelaksanaan proses akreditasi program seperti pelan strategik universiti yang tidak selari dengan proses akreditasi program, kekurangan penglibatan pihak berkepentingan, beban kerja dan birokrasi. Walaupun begitu, terdapat perubahan positif bagi reka bentuk kurikulum, sistem sokongan pelajar, dan budaya kualiti dengan pelaksanaan akreditasi program.

Menurut Emmanuel Boakye-Yiadom (2021), penilaian memainkan peranan yang penting dalam membantu memperbaiki dan meningkatkan aktiviti yang dijalankan di universiti. Individu, keluarga, persekitaran kerja, program pengajian dan institusi menjadi faktor kepada kejayaan usaha yang dijalankan. Dalam bidang pendidikan, penilaian memainkan fungsi yang penting untuk memastikan proses pengajaran yang lancar dan berkesan.

Kebanyakan kajian yang dijalankan di atas ialah sorotan kajian berkenaan prestasi graduan dan prestasi institusi. Namun bagi kajian prestasi yang ingin dilaksanakan ini, ia merujuk kepada faktor yang ada yang mempengaruhi pencapaian penilaian program akreditasi baharu. Faktor ini boleh dibahagikan kepada tiga iaitu sumber manusia, permohonan dan program. Sumber manusia terdiri daripada bahagian yang mengurus setiap proses penilaian dan panel penilai yang membuat penilaian ke atas program. Permohonan merujuk kepada pihak yang memohon, kategori pemohon, taraf pemohon dan negeri pemohon, manakala program pula merujuk kepada kategori, tahap, peringkat, status dan bidang program. Faktor-faktor ini dikenal pasti bagi meramal pencapaian prestasi di masa hadapan sama ada penilaian dapat dilaksanakan di dalam atau di luar tempoh yang ditetapkan.

Melalui sains data, analisis dapat dijalankan bagi mengenal pasti petunjuk atau memberi *insights* tentang faktor-faktor yang mempengaruhi prestasi manakala algoritma pembelajaran mesin pula dilatih menggunakan data sejarah untuk membina model ramalan yang boleh meramal prestasi masa hadapan.

2.2.3 Kaedah Analisis Prestasi

Ramalan prestasi merujuk kepada analisis dengan menggunakan data yang ada untuk membuat ramalan tentang prestasi masa hadapan individu atau sistem. Ia digunakan untuk membuat keputusan dan perancangan masa hadapan dalam pelbagai bidang seperti kewangan, akademik, sukan dan sumber manusia. Antara kaedah yang sering digunakan dalam pelbagai bidang untuk meramal prestasi ialah kaedah statistik dan kaedah pembelajaran mesin (Prof. Avijit Mondal 2023).

a. Kaedah statistik

Beberapa kaedah statistik digunakan untuk menganalisis prestasi. Assous (2022) dan Abduh (2014) menggunakan analisis regresi untuk mencari hubungan antara pelbagai pemboleh ubah tidak bersandar iaitu faktor-faktor yang mempengaruhi kecekapan dengan pemboleh ubah bersandar dalam meramal prestasi dan kecekapan perbankan.

Selain itu, kaedah yang biasa digunakan untuk menilai kecekapan ialah Analisis Penyampulan Data (*DEA*). Mohamad Razi et al. (2022) menggunakan model DEA untuk mengira skor kecekapan setiap pelajar berdasarkan nisbah output kepada input. Althebyan et al. (2016) pula menggunakan model DEA bagi mengukur prestasi institusi pengajian tinggi khususnya di Jordan bagi menilai kecekapan universiti berdasarkan pelbagai pemboleh ubah input dan output. DEA sesuai digunakan untuk melihat interaksi antara berbilang input (faktor pengaruh) dan output (prestasi) dan digunakan sebagai penanda aras prestasi.

b. Kaedah pembelajaran mesin

Selain kaedah statistik, kaedah pembelajaran mesin digunakan oleh penyelidik bagi menerangkan hubungan di antara faktor pengaruh dengan prestasi. Kaedah ini digunakan untuk mencipta model yang meramal masa hadapan dengan tepat berdasarkan set pemboleh ubah atau ciri input. (*Types of Predictive Analytics Models and Their Applications*, 2023). Beberapa kaedah pembelajaran mesin yang digunakan dalam menganalisis prestasi ialah seperti pengelasan dan regresi yang merupakan

contoh kaedah pembelajaran mesin terselia manakala kluster dan pengecaman imej ialah contoh kaedah pembelajaran mesin tidak terselia.

Ramai penyelidik menganalisis prestasi pelajar di institusi pendidikan tinggi menggunakan kaedah pengelasan. Bagi mengenal pasti pelajar yang berisiko gagal, Patel (2021) dan Hashim et al. (2020) menggunakan teknik Pokok Keputusan PK, Naive Bayes NB, Rangkaian Saraf Buatan RSB dan Regresi Logistik RL untuk membina model yang boleh meramal prestasi akademik pelajar berdasarkan faktor pencapaian akademik dan demografi. Kajian Suwitno & Wibowo (2019) dan Yuliansyah et al. (2021) pula menggunakan C4.5 untuk meramal pelajar yang bergraduasi pada masa yang ditetapkan. Selain itu, k-Jiran Terdekat k-JT juga digunakan untuk meramal prestasi pelajar seperti yang digunakan di dalam kajian Hashim et al. (2020) dan Yusof et al. (2022).

Melalui kajian Mohamed Nafuri et al. (2022), kluster digunakan bagi mengelaskan pelajar B40 berdasarkan prestasi, sikap dan pencapaian mereka di IPT awam di Malaysia. Teknik kluster digunakan untuk mengumpulkan beberapa objek yang serupa ke dalam satu kelompok yang sama. Teknik ini boleh digunakan bagi set data yang tiada label atau kategori.

2.3 PEMAHAMAN PENILAIAN PROGRAM AKREDITASI

Akreditasi program ialah jaminan ke atas kualiti program yang ditawarkan oleh PPT bagi memastikan program yang ditawarkan mencapai kualiti yang diperlukan dengan mengikut kod amalan dan standard yang ditetapkan.

Akreditasi digunakan sebagai indikator kualiti. Akreditasi boleh dikategorikan pada peringkat institusi dan peringkat program. Akreditasi diperingkat institusi ialah dengan menyemak keseluruhan proses dan kualiti sesebuah institusi, manakala akreditasi diperingkat program pula ialah menyemak program yang dijalankan oleh institusi. Hasil daripada status akreditasi mempunyai impak yang besar terhadap banyak aspek di PPT. Kajian Kumar et al. (2021) menunjukkan kesan akreditasi ke atas pendidikan tinggi. Ia membantu untuk meningkatkan kualiti pendidikan tinggi melalui penambahbaikan dasar, proses dan bidang fungsi terasnya, seperti penyelidikan, akademik, pengajaran-

pembelajaran dan sebagainya. Pelaksanaan akreditasi juga dapat mempertingkatkan hasil pembelajaran dan kebolehpasaran graduan (Parscale et al. 2022). Akreditasi amat penting supaya orang ramai dapat membuat pemilihan institusi dengan tepat (Noripansyah et al. 2023).

Proses akreditasi pada mulanya adalah proses yang dilaksanakan secara sukarela, namun apabila objektifnya berkembang terhadap peraturan awam, prosedur mendapatkan akreditasi ini menjadi wajib. Ia bukanlah proses yang hanya dijalankan sekali sahaja, namun menurut Giraud (2001) ia adalah proses peningkatan kualiti berterusan. Menurut kajian Musthafa et al. (2020) penilaian akreditasi melibatkan masa, sumber manusia dan kewangan.

2.4 PERLOMBONGAN DATA

Perlombongan data adalah proses mendapatkan maklumat daripada set data. Istilah perlombongan data kebanyakannya digunakan oleh komuniti statistik, penganalisis data, dan sistem maklumat pengurusan. Ia juga mendapat populariti dalam bidang pangkalan data (Fayyad et al. 1996). Set data yang besar tetap dikira sebagai data sehingga diekstrak maklumat daripada data tersebut untuk mendapatkan maklumat berguna daripadanya (Dhawan 2018). Perlombongan data ialah kesinambungan daripada analisis data tradisional dan pendekatan statistik kerana ia menggabungkan teknik analisis yang diambil daripada pelbagai disiplin dan tidak terhad kepada analisis nombor, padanan corak dan bidang kecerdasan buatan seperti pembelajaran mesin, rangkaian saraf dan algoritma genetik (Jackson 2002).

2.4.1 Pemilihan Fitur

Antara aktiviti peringkat awal perlombongan data setelah pembersihan data ialah membuat pemilihan fitur. Pemilihan fitur ialah proses memilih subset ciri yang relevan dan penting (M. Kumar et al. 2022) bertujuan untuk meningkatkan prestasi model (Huang 2015).

Fitur yang banyak menyebabkan bilangan dimensi bertambah dan mengakibatkan titik data menjadi lebih terpicil dalam ruang ciri, menjadikannya lebih

sukar untuk mencari corak dan hubungan. Menurut Farissi et al. (2020), pemilihan ciri bertujuan untuk mengatasi set data berdimensi tinggi dengan mengurangkan dimensi fitur dalam set data. Pengurangan dimensi ini bertujuan meningkatkan prestasi pembelajaran mesin.

Melihat kepada impak pemilihan fitur ke atas prestasi model, bidang ini semakin mendapat perhatian dan semakin berkembang dikalangan pengkaji di dalam perlombongan data. Terdapat pengkaji yang menggunakan pelbagai kaedah pemilihan fitur dalam kajian mereka seperti Zaffar et al. (2021) dan Hamoud & Humadi (2019) . Terdapat juga kajian yang memeriksa kelebihan (Saifudin et al. 2020) dan limitasi (Pirgazi et al. 2019) dalam pemilihan fitur.

Beberapa kaedah boleh digunakan untuk membuat pemilihan fitur seperti kaedah penapis, kaedah pembalut dan kaedah hibrid. Antara contoh kaedah penapis yang boleh digunakan ialah teknik *Information Gain* (IG), Ujian *Chi-square* dan analisis korelasi. Contoh teknik yang digunakan bagi kaedah pembalut pula ialah Pemilihan Berjjukan ke Depan (PBD), Penyingkiran Berjjukan ke Belakang (SBB), Penyingkiran Ciri Rekursif (PCR) dan carian rawak Algoritma Genetik (AG) manakala contoh kaedah hibrid ialah penapis-pembalut hibrid.

Mengikut kajian, kaedah penapis boleh digunakan untuk tujuan perlombongan data mengikut tugas sama ada pengelasan, regresi atau pengelompokan (*clustering*). Bagi teknik pengelasan, IG, *Gain Ratio* dan *Chi Square* adalah teknik yang sesuai (Jović et al. 2015). Kaedah penapis dikatakan lebih cepat kerana tidak menggunakan algoritma pembelajaran dan sesuai untuk data berdimensi tinggi tetapi kurang ketepatan dalam memilih fitur (Pirgazi et al. 2019).

Kaedah pembalut pula menggunakan algoritma pembelajaran mesin untuk menghasilkan subset data bagi membangunkan model. Model yang dibangunkan pula akan menilai ciri-ciri data berdasarkan skor yang diperolehi. Kaedah pembalut dikatakan agak perlahan berbanding penapis kerana melaksanakan teknik pengulangan dalam mencari subset ciri yang terbaik. Kaedah ini boleh menghasilkan prestasi yang

baik sekiranya kurang data kotor, bilangan fitur sedikit dan berkaitan serta model pembelajaran mesin yang sesuai dipilih (S. H. Huang 2015).

Kaedah hibrid menggabungkan kelebihan kaedah penapis dan kaedah pembalut untuk mencapai keseimbangan antara kecekapan dan keberkesanan dalam pemilihan ciri khususnya bagi set data berdimensi tinggi.

Walaupun pelbagai kaedah pemilihan fitur boleh digunakan, ketiga-tiga kaedah mempunyai kekangan atau had seperti dipengaruhi oleh saiz data (*scalability*), keperluan algoritma pembelajaran mesin (*stability*) dan pemilihan parameter. Kaedah pemilihan fitur bagi tugas pengelasan hendaklah melihat kepada sumber yang ada, matlamat yang ingin dicapai, teknik perlombongan yang digunakan dan tahap pengoptimuman yang dikehendaki (Jović et al. 2015). Bagi mengetahui dengan lebih lanjut berkenaan ketiga-tiga kaedah pemilihan fitur ini, ia akan diterangkan dengan lebih terperinci di dalam subtopik di bawah.

a. Kaedah penapis

Pemilihan fitur kaedah penapis menggunakan teknik statistik untuk menilai hubungan antara setiap pemboleh ubah input dan pemboleh ubah sasaran. Pemilihan fitur adalah bebas daripada sebarang algoritma pembelajaran mesin sebaliknya fitur dipilih berdasarkan skor daripada ujian statistik. Skor yang diperolehi akan digunakan sebagai asas untuk memilih pemboleh ubah input yang akan digunakan dalam model.

Jenis input dan output akan menentukan ujian statistik yang bersesuaian untuk pemilihan fitur contohnya IG dan *Chi-square* sesuai dengan input dan output jenis nominal, pekali korelasi seperti *Pearson's Correlation Coefficient* pula sesuai dengan data numerik manakala ANOVA sesuai dengan data numerik dan nominal.

Kaedah penapis adalah lebih pantas berbanding kaedah pembalut kerana ia tidak melibatkan latihan model. Kaedah ini digunakan oleh Pattiasina & Rosiyadi (2020) dan Yusof et al. (2022) untuk mencari subset fitur yang paling optimum dengan model ramalan yang dibangunkan. Ia sesuai digunakan untuk memilih subset fitur dari set data yang besar namun kaedah penapis mungkin gagal untuk mencari subset fitur terbaik

dalam banyak keadaan kerana ia menilai fitur secara bebas antara satu sama lain. Oleh itu ia mungkin mengabaikan gabungan fitur yang mempunyai maklumat berharga yang saling berkait.

b. Kaedah pambalut

Kaedah pambalut ialah kaedah pemilihan fitur dengan membuat penilaian kepada beberapa prosedur menambah atau menyingkirkan fitur untuk mencari gabungan optimum yang memaksimumkan prestasi model. Model yang berprestasi terbaik dengan subset fitur optimum akan dipilih. Kaedah pambalut menggunakan algoritma seperti HR, PK dan NB untuk menjana subset fitur. Walaupun berkesan dalam menyingkir fitur yang tidak relevan, ia adalah lebih perlahan kerana menggunakan algoritma perlombongan berkali-kali. Ia akan menukar bilangan fitur pada setiap satu pelaksanaan carian kerana mengikut kriteria carian dan berhenti (Mejía-Lavalle & Morales t.th.). Contoh pemilihan fitur kaedah pambalut ialah PBD, SBB dan PCR.

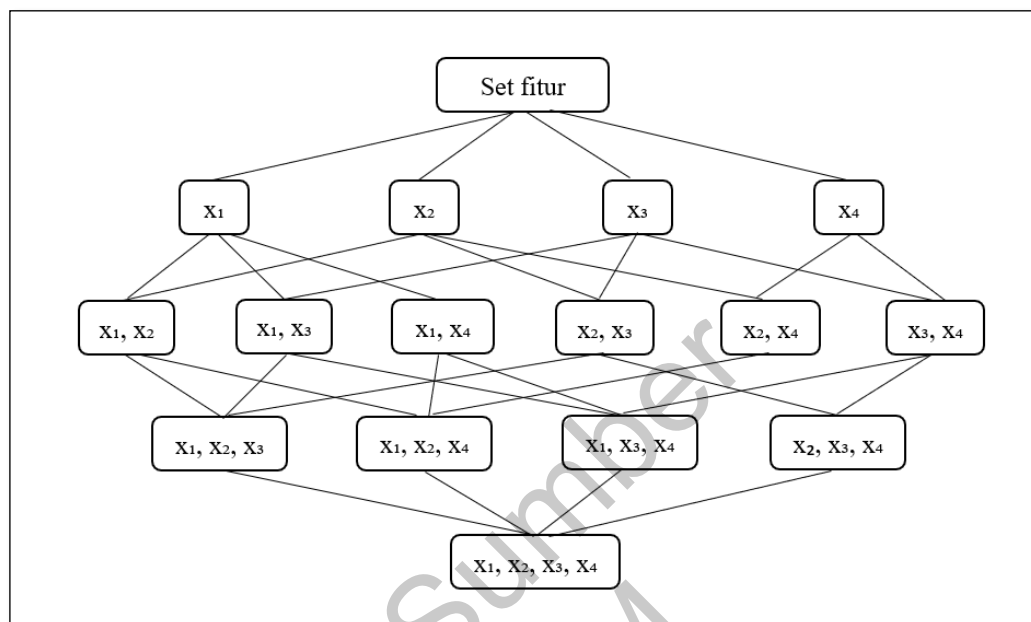
Strategi carian untuk menghasilkan subset fitur ialah aspek teras kaedah pambalut. Ia merujuk kepada proses meneroka ruang kombinasi fitur yang mungkin dan menilai keberkesanannya dalam model pembelajaran mesin yang dipilih. Terdapat beberapa strategi carian bagi pemilihan fitur kaedah pambalut iaitu strategi carian menyeluruh, strategi carian berjujukan dan strategi carian rawak seperti yang akan dihuraikan di bawah.

i. Strategi carian menyeluruh

Strategi ini menilai semua subset fitur yang mungkin bagi mencari penyelesaian yang optimum. Carian menyeluruh mencipta semua subset yang mungkin bagi set fitur sebenar dan mengesyorkan subset terbaik mengikut kriteria. Strategi carian ini mengambil masa yang lama kerana akan menilai kombinasi $2^p - 1$ untuk bilangan p fitur yang diperlukan.

Sebagai contoh, terdapat empat fitur iaitu $\{x_1, x_2, x_3, x_4\}$. Oleh itu, jumlah fitur yang akan dihasilkan dan dinilai ialah $2^4 - 1 = 15$. Proses menghasilkan fitur dan penilaian adalah bermula dengan $\{x_1\}$, $\{x_2\}$... $\{x_4\}$. Kemudian subset dengan dua fitur

dihasilkan seperti $\{x_1, x_2\}$, $\{x_1, x_3\}$... $\{x_2, x_4\}$ dan seterusnya ulangan proses sehingga semua fitur dipilih. Rajah 2.1 menunjukkan proses kaedah pembalut strategi carian menyeluruh menyeluruh menghasilkan subset fitur.



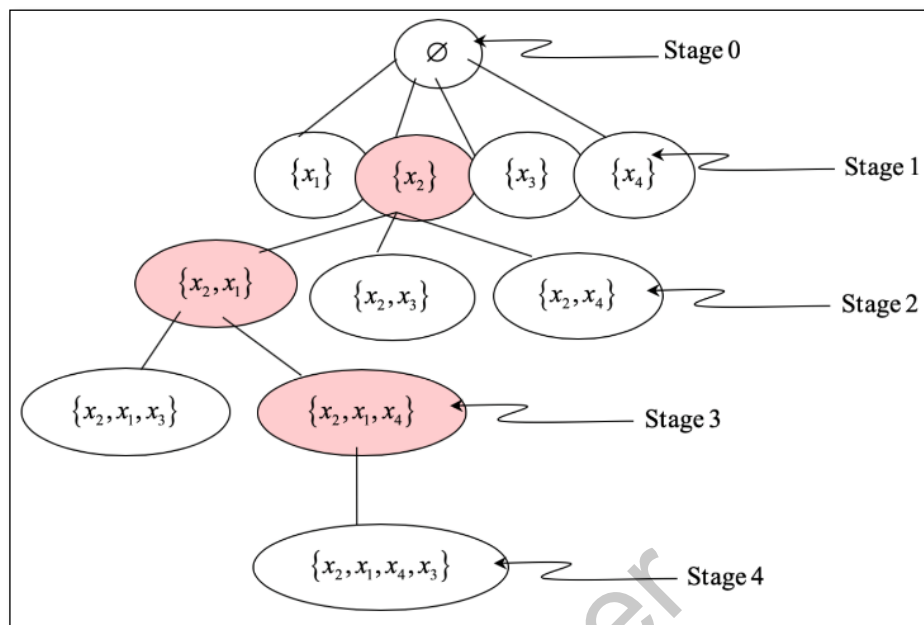
Rajah 2.1 Proses menghasilkan subset fitur strategi carian menyeluruh

Selain mengambil masa yang lama, kos strategi carian menyeluruh juga adalah mahal dan tidak sesuai untuk set data yang besar.

ii. Strategi carian berjujukan

Strategi carian berjujukan menggunakan pelbagai teknik untuk meneroka ruang carian dengan cekap. Beberapa teknik carian berjujukan ialah seperti Pemilihan Berjujukan ke Depan PBD, Penyingkiran Berjujukan ke Belakang SBB dan Berjujukan Dua Arah BDA.

PBD akan memulakan proses carian tanpa set fitur dalam model dan secara berulang menambah satu fitur pada satu masa sehingga kriteria berhenti dipenuhi atau sehingga tiada peningkatan prestasi berlaku. Rajah 2.2 menunjukkan contoh proses pemilihan fitur strategi carian PBD.



Rajah 2.2 Proses pemilihan fitur strategi PBD

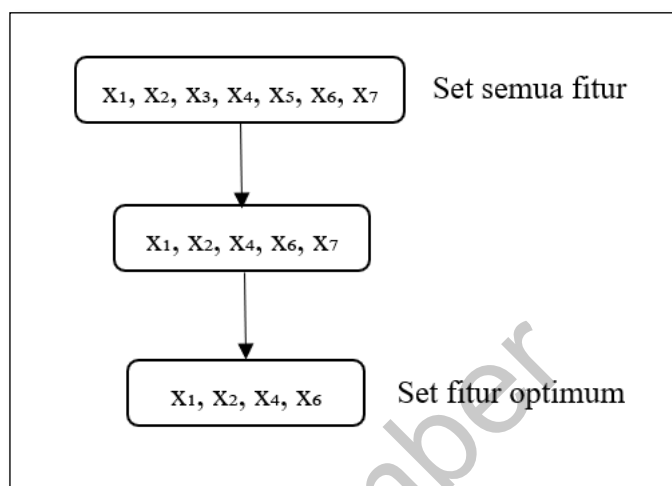
Sumber: Qifan He dan Dr. Rajen Bhatt 2020

Rajah 2.2 menerangkan contoh proses carian PBD dengan empat fitur dalam set asal iaitu $\{x_1, x_2, x_3, x_4\}$. PBD bermula dengan fitur kosong, kemudian menilai setiap model fitur mengikut kriteria yang ditetapkan dan fitur yang terbaik dikekalkan. Setelah itu, dua fitur yang terbaik yang dipilih akan digabungkan. Proses ini diulang sehingga tiada peningkatan prestasi atau semua fitur telah digunakan. Contohnya pada tahap 3, set fitur $\{x_2, x_1, x_4\}$ terpilih kerana prestasi yang sama atau lebih baik dengan set fitur $\{x_2, x_1, x_4, x_3\}$.

Kajian Fahrudy & Uyun (2022) menggunakan strategi carian PBD. Kajian ini membandingkan pemilihan fitur PBD dengan IG dan dapatan kajian ialah pemilihan ciri kaedah pembalut dengan strategi carian PBD dapat memberi prestasi yang lebih baik dari segi ketepatan pengelasan. Kekurangan utama PBD ialah ia tidak dapat mengeluarkan fitur yang dikekalkan setelah penambahan fitur-fitur lain.

Strategi carian SBB pula berlawanan dengan PBD. Strategi carian ini bermula dengan semua set fitur digunakan dan akan menyingkirkan satu demi satu fitur yang mempunyai kesan paling sedikit pada prestasi model secara berulang sehingga kriteria berhenti dipenuhi. Sirodjudin et al. (2016) menggunakan kaedah pembalut SBB bagi

membuat pemilihan fitur yang paling memberi pengaruh ke atas ramalan mutu dan kualiti beras. Rajah 2.3 menunjukkan ilustrasi proses pemilihan fitur dengan strategi SBB.

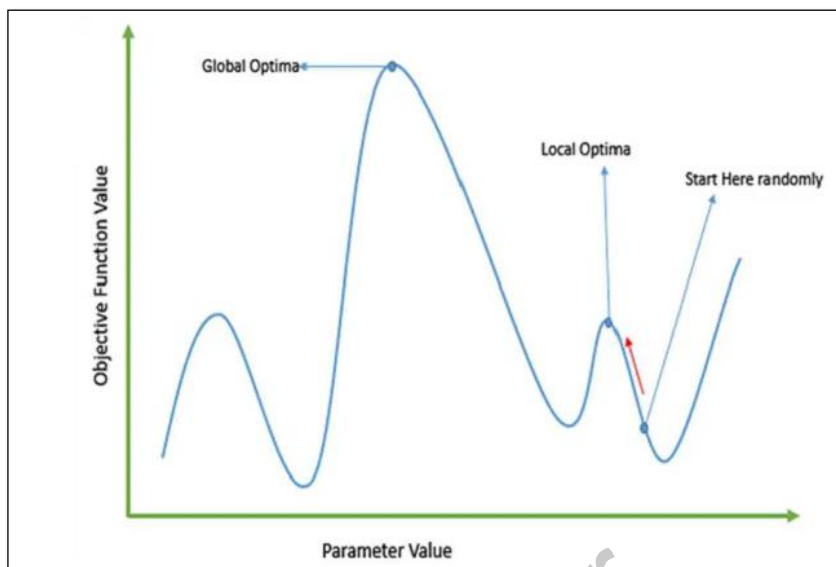


Rajah 2.3 Proses pemilihan fitur strategi SBB

Kelebihan SBB ialah ia berfungsi paling baik apabila subset fitur optimum adalah besar kerana SBB akan memulakan carian dengan semua fitur. Kekurangan SBB yang utama ialah ketidakupayaannya untuk menilai semula kegunaan sesuatu fitur yang telah dibuang.

Strategi carian berjjukan yang seterusnya ialah BDA. Pelaksanaan carian melalui strategi ini ialah cariannya dilaksanakan serentak diantara PBD dan SBB. PBD dimulakan dari set fitur kosong manakala SBB dimulakan dari set semua fitur. Untuk memastikan bahawa PBD dan SBB menumpu kepada penyelesaian yang sama, fitur yang telah dipilih oleh PBD tidak akan dibuang oleh SBB dan fitur yang telah dibuang oleh SBB tidak dipilih semula oleh PBD.

Kaedah menghasilkan subset fitur melalui strategi carian PBD, SBB dan BDA boleh menyebabkan isu *local optima*. *Local optima* terjadi apabila subset fitur yang dihasilkan adalah optima pada tahap lokal sahaja dan bukan kepada subset fitur *global optima* seperti yang ditunjukkan pada Rajah 2.4.



Rajah 2.4 Ilustrasi lokal optima dan global optima
Sumber: Sam Mills 2018

iii. Strategi carian rawak

Strategi carian rawak ialah strategi di mana subset fitur dipilih secara rawak dan dinilai untuk mencari set fitur yang optimum. Salah satu contoh strategi carian rawak ialah AG. Strategi carian rawak AG menggunakan prinsip evolusi berdasarkan analogi pemilihan semula jadi dan genetik Darwinian dalam sistem biologi dan dikaitkan dengan ciri-ciri pencarian eksplorasi dan eksploitasi.

Menurut C. L. Huang & Wang (2006) kelebihan carian rawak AG ialah menggunakan ruang carian yang besar dengan cekap dan tidak menghadapi masalah pada *local optima* berbanding strategi carian berjjukan.

c. Kaedah hibrid

Kaedah hibrid ialah kaedah pemilihan fitur dengan membuat penilaian menggabungkan pelbagai teknik pemilihan ciri untuk mencari pendekatan yang lebih berkesan dalam konteks pemilihan fitur (Pirgazi et al. 2019).

Kaedah hibrid bertujuan memanfaatkan kekuatan kaedah yang berbeza seperti kaedah penapis-pembalut hibrid. Pada peringkat penapis, fitur yang tidak relevan dikeluarkan dengan cepat berdasarkan ukuran statistik atau teori maklumat. Pada

peringkat pembalut pula subset fitur yang terpilih diperhalusi dengan mengambil kira interaksinya dengan prestasi model.

2.4.2 Teknik Pemilihan Fitur Dalam Kajian Ramalan Prestasi Pendidikan Tinggi

Terdapat banyak teknik pemilihan fitur yang digunakan di dalam kajian ramalan prestasi. Kajian Mengash (2020) menggunakan teknik regresi linear untuk mencari hubungan diantara pemboleh ubah tidak bersandar iaitu gred purata sekolah tinggi (HSGA), pencapaian skor ujian kemasukan *scholastic* (SAAT) dan skor ujian bakat/kebolehan (GAT) dengan pemboleh ubah bersandar iaitu gred purata markah pelajar bagi dua (2) semester awal pengajian dan menggunakan pekali korelasi untuk melihat kekuatan hubungan diantara pemboleh ubah tersebut. Berdasarkan ujian awal kemasukan ini mendapati pencapaian skor ujian kemasukan SAAT merupakan kriteria yang paling tepat meramal pencapaian pelajar setelah melanjutkan pelajaran ke peringkat universiti.

Kajian M. Kumar et al. (2022) menggunakan analisis korelasi, IG dan *gain ratio* untuk memilih ciri yang paling relevan bagi meramal prestasi akademik pelajar menggunakan pelbagai algoritma pengelasan.

Bagi memilih fitur yang mempengaruhi pengekalan pelajar di IPT, Febro (2019) menggunakan kaedah penapis iaitu analisis korelasi, IG dan *Chi-square*. Pengekalan pelajar di IPT dikaitkan secara positif dengan fitur markah ujian kemasukan ke IPT, jumlah adik-beradik, gred Bahasa Inggeris dan Matematik dan juga pendapatan keluarga.

Demografi, latar belakang akademik dan sikap pelajar merupakan faktor yang boleh mempengaruhi prestasi akademik pelajar. Farissi et al. (2020) dalam kajiannya menggunakan carian rawak AG bagi memilih fitur tersebut bersama algoritma pembelajaran mesin bagi meningkatkan ketepatan ramalan prestasi akademik pelajar.

Bagi mengenal pasti faktor paling relevan yang mempengaruhi prestasi akademik pelajar dan meningkatkan ketepatan ramalan model, Saifudin et al. (2020) menggunakan kaedah PBD untuk memilih atribut yang memenuhi kriteria. Melalui

kajian tersebut, apabila bilangan fitur yang dipilih meningkat, prestasi model ramalan semakin berkurangan. Ini bermakna bahawa banyak fitur tidak relevan atau menyebabkan bias. Pemilihan hanya beberapa fitur telah meningkatkan prestasi ramalan model pengelasan.

Terdapat juga pengkaji yang menggunakan lebih daripada satu kaedah bagi memilih fitur. Kajian Zaffar et al. (2021) menggunakan teknik hibrid iaitu gabungan ciri penapis-pemalut. Melalui kajian ini, kaedah pemilihan ciri hibrid bukan sahaja meningkatkan ketepatan ramalan, tetapi juga digunakan untuk mengenal pasti ciri optimum bagi merancang strategi produktif untuk meningkatkan prestasi akademik pelajar.

Jadual 2.1 merupakan teknik pemilihan fitur yang digunakan di dalam beberapa kajian bagi ramalan prestasi pendidikan tinggi.

Jadual 2.1 Teknik pemilihan fitur untuk ramalan prestasi pendidikan tinggi

| Pengarang & Tahun | Objektif | Teknik |
|-------------------------------|--|--|
| M. Kumar et al. (2022) | Mengenal pasti fitur yang paling mempengaruhi prestasi pelajar. | Kaedah penapis IG, <i>Gain Ratio</i> dan kolerasi |
| Fahrudy & Uyun (2022) | Mengenal pasti fitur yang paling mempengaruhi untuk meramal pelajar tamat belajar dan membandingkan prestasi dua kaedah pemilihan fitur. | Kaedah penapis IG dan kaedah pemalut PBD |
| Bedregal-Alpaca et al. (2020) | Mengenal pasti pelajar yang berisiko tercicir berdasarkan prestasi akademik, ciri peribadi dan faktor lain yang berkaitan. | Kaedah penapis IG dan kolerasi dan kaedah pemalut PCR |
| Yusof et al. (2022) | Mengenal pasti ciri mempengaruhi prestasi akademik pelajar yang mempunyai latar belakang dan ciri pembelajaran yang berbeza. | Kaedah penapis IG, <i>Symmetrical Uncertainty</i> , <i>Gain Ratio</i> dan kolerasi |
| Farissi et al. (2020) | Menggabungkan AG dengan pengelas HR untuk meningkatkan ketepatan ramalan prestasi akademik pelajar. | Carian rawak AG |
| Khasanah & Harwati (2017) | Mengenal pasti fitur yang mempengaruhi sama ada pelajar akan tercicir atau tidak. | Kaedah penapis kolerasi |
| Saifudin et al. (2020) | Mengenal pasti faktor paling mempengaruhi prestasi akademik pelajar bagi meningkatkan ketepatan ramalan model. | Kaedah pemalut PBD |
| Zaffar et al. (2021) | Membangunkan rangka kerja pemilihan ciri hibrid untuk mengenal pasti fitur yang | Kaedah hibrid penapis-pemalut bersambung... |

| | | |
|------------------------|--|--|
| ...sambungan | paling mempengaruhi ramalan prestasi pelajar. | PBD |
| Hamoud & Humadi (2019) | Untuk menentukan kedudukan dan memilih 30 soalan soal selidik teratas daripada 61 soalan yang mempengaruhi kejayaan pelajar. | Kaedah penapis IG, kolerasi dan <i>SVM Attribute Eval</i> |
| Febro (2019) | Menganalisis faktor yang mempengaruhi kejayaan pelajar dalam kalangan pelajar tahun satu universiti melalui pemilihan fitur. | Kaedah penapis IG, kolerasi dan analisis <i>chi-square</i> |

Teknik-teknik pemilihan fitur yang dinyatakan di atas adalah daripada pelbagai kajian prestasi di dalam sektor pendidikan tinggi. Namun belum ada kajian menggunakan teknik pemilihan fitur bagi memilih subset fitur yang mempengaruhi prestasi penilaian program akreditasi baharu. Oleh itu, kajian ini akan membandingkan lima teknik pemilihan fitur iaitu (1) teknik penapis IG, (2) teknik pembalut PBD, (3) teknik pembalut SBB, (4) teknik carian rawak AG dan (5) kaedah hibrid teknik penapis-pembalut PBD.

2.4.3 Pengelasan

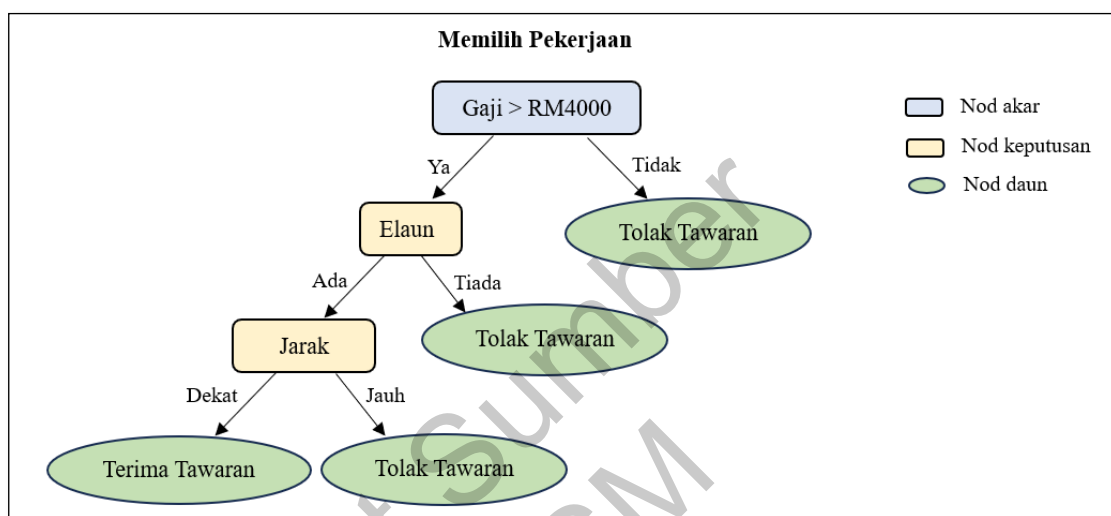
Pengelasan merupakan salah satu teknik perlombongan data bagi meramal kelas sasaran sesuatu data. Algoritma pengelasan digunakan untuk meramal data secara automatik ke dalam kategori atau kelas berdasarkan cirinya. Terdapat pelbagai jenis algoritma pengelasan, masing-masing mempunyai kegunaan, kekuatan dan kelemahannya sendiri. Algoritma terbaik adalah bergantung kepada sifat data dan hasil yang diinginkan. Beberapa teknik pengelasan yang popular dan banyak digunakan di dalam kajian dinyatakan di dalam subtopik seterusnya.

a. Pokok Keputusan

Pokok Keputusan (PK) membina pengelasan dalam bentuk struktur pokok. Ia mengelaskan data dengan menyusunnya berdasarkan nilai-nilai ciri dengan memecah satu set data ke subset yang lebih kecil sementara pada masa yang sama pokok keputusan yang dikaitkan semakin berkembang. Lebih terperinci pokok itu, lebih kompleks peraturan keputusan dan cenderung menghadapi masalah terlebih padanan.

Bagi mengatasi masalah ini, kaedah pemangkasan pokok dilakukan bagi menjadikannya lebih mudah untuk difahami (Patil et al. 2010).

Rajah 2.5 menunjukkan ilustrasi contoh PK. Setiap nod di dalam PK menerangkan ciri data yang akan dikelaskan. Setiap ranting menunjukkan keputusan ujian dan setiap nod daun mewakili pengelasan kelas sasaran.



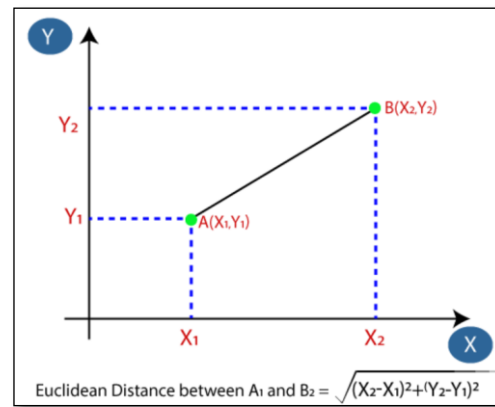
Rajah 2.5 Contoh Model Pengelasan PK

PK sangat popular kerana ia menghasilkan peraturan pengelasan yang mudah ditafsir berbanding teknik pengelasan lain (Kumar Yadav 2012).

b. K-Jiran Terdekat

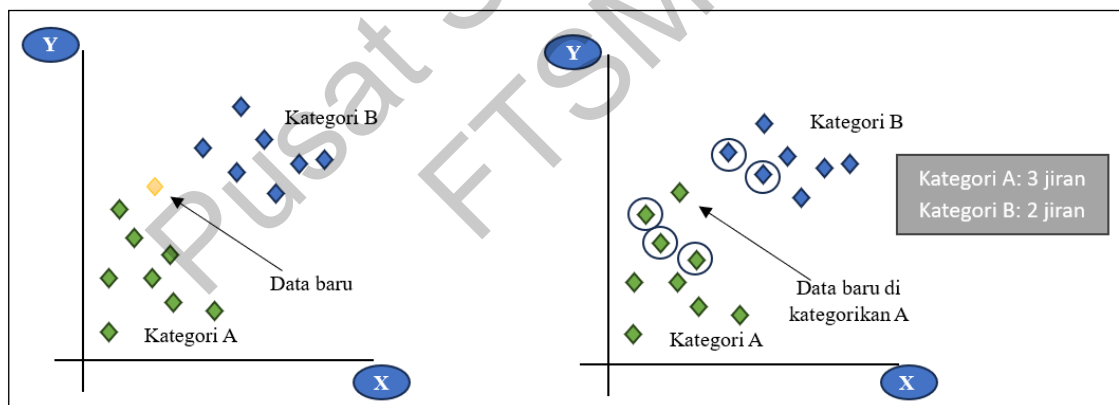
Algoritma k-Jiran Terdekat (k-JT) ialah kaedah untuk mengelaskan objek berdasarkan data pembelajaran yang paling hampir dengan objek tersebut. Ia digunakan untuk mengelaskan data tanpa label dengan meletakkannya kepada kelas label yang paling hampir atau sama dengan mengira jarak dari semua titik dalam lingkungan data yang tidak diketahui dan memilih data yang mempunyai jarak terdekat dengannya.

Jiran Terdekat ialah pendekatan untuk mengira jarak antara kes baharu dan kes lama berdasarkan padanan beberapa ciri. Secara umumnya, formula *Euclidean* digunakan untuk menentukan jarak antara dua titik iaitu jarak A dan B seperti Rajah 2.6.

Rajah 2.6 Formula *Euclidean*

Sumber: *K-Nearest Neighbor (KNN) Algorithm for Machine Learning*

Rajah 2.7 di bawah menunjukkan bagaimana k-JT berfungsi. Contohnya, tetapkan jumlah k-jiran kepada lima (5). Dengan mengira jarak *Euclidean*, didapati tiga (3) jiran terdekat dalam kategori A dan dua (2) jiran terdekat dalam kategori B. Oleh itu, data baharu termasuk di dalam kategori A.



Rajah 2.7 Contoh Model Pengelasan k-JT

k-JT ialah model pengelasan yang mudah dilaksanakan pada set data yang kecil, tetapi ia memberikan prestasi yang lebih perlahan apabila digunakan pada data yang banyak dan berdimensi tinggi (Ranaware & Potdar, 2016).

c. Naïve Bayes

Algoritma Naïve Bayes (NB) merupakan model pengelasan yang menggunakan Teori *Bayes (Bayes Theorem)* yang telah diperkenalkan oleh *Thomas Bayes*. Teori *Bayes*

adalah model berdasarkan statistik yang mengira kebarangkalian. Kelas dengan kebarangkalian tertinggi dianggap sebagai kelas sasaran yang paling mungkin. Algoritma ini berfungsi pada konsep yang mudah dan merupakan model yang mempunyai ciri bebas dengan menganggap bahawa fitur-fitur adalah tidak bersandar antara satu sama lain (Karthika & Sairam, 2015).

Formula Teori *Bayes* $P(A|B)$ ialah kebarangkalian:

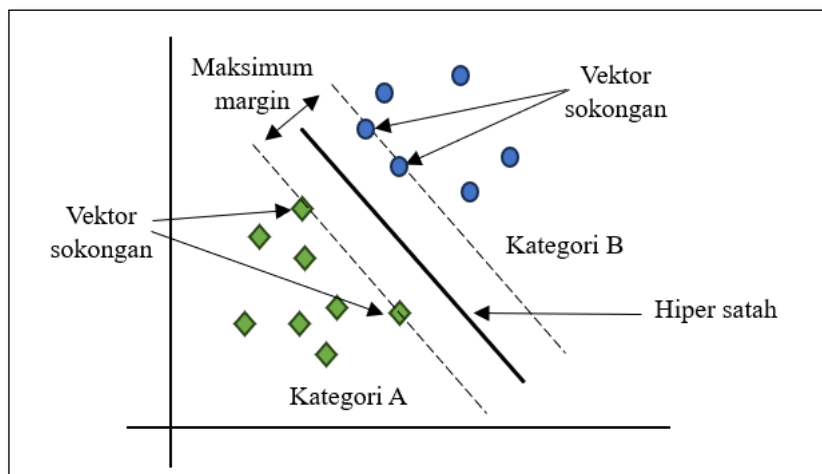
$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad \dots(2.1)$$

NB ringkas, mudah untuk diimplementasi dan pantas. Walaupun ia mudah, namun ia kerap mengatasi teknik pengelasan yang lain (Kesarwani & Tiwari 2018). Selain itu, algoritma ini tidak terkesan kepada data hingar (Shaziya et al. 2007).

d. Mesin Sokongan Vektor

Mesin Sokongan Vektor (MSV) adalah salah satu algoritma pembelajaran mesin yang diselia (*supervised*) yang sering digunakan dalam model pengelasan bagi mengenal pasti penyelesaian yang optimum. MSV adalah berdasarkan konsep pesawat yang menentukan sempadan dengan memisahkan antara satu set objek yang terdiri daripada kelas yang berbeza.

Matlamat algoritma MSV adalah untuk mencipta garisan terbaik yang boleh mengasingkan ruang n-dimensi ke dalam kelas supaya boleh meletakkan titik data baharu dengan mudah dalam kategori yang betul pada masa hadapan. Sempadan keputusan terbaik ini dipanggil hiper satah. Vektor terhampir bagi kedua-dua kelas dengan hiper satah dipanggil vektor sokongan. Jarak antara vektor sokongan dengan hiper satah dipanggil margin. Saiz margin yang maksimum dapat menghasilkan model pengelasan yang lebih baik seperti Rajah 2.8.



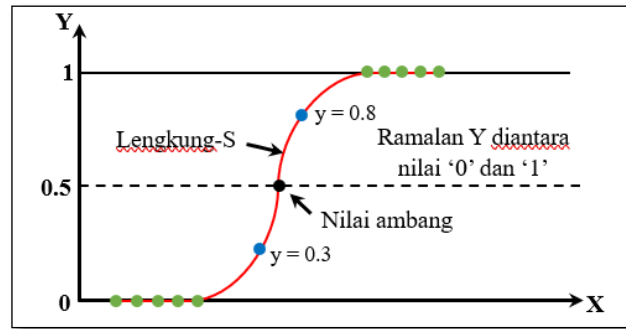
Rajah 2.8 Ilustrasi MSV

Dengan pemahaman yang lebih mudah dan ringkas, MSV mencari sempadan terbaik dalam mengasingkan kelas. Kelebihan utama MSV ialah ia sangat efektif bagi ruang berdimensi tinggi (Soofi & Awan 2017).

e. Regresi Logistik

Algoritma Regresi Logistik (RL) adalah salah satu algoritma pembelajaran mesin yang paling biasa digunakan untuk pengelasan binari dengan contoh kelas sasaran 'Ya' atau 'Tidak', 'Betul' atau 'Salah' dan '0' atau '1'. Matlamat utama RL adalah untuk mencari hubungan diantara pemboleh ubah bersandar (kelas sasaran) dengan pemboleh ubah tidak bersandar dengan menganggar kebarangkalian kejadian (Kumar Verma & Lokesh Bewariya 2021).

Nilai RL adalah di antara 0 dan 1 menjadikan ia berbentuk lengkung seperti bentuk 'S'. Lengkung bentuk-S dipanggil fungsi *Sigmoid* atau fungsi logistik. Nilai ambang digunakan untuk mentakrif kebarangkalian sama ada '0' atau '1'. Seperti Rajah 2.9, nilai di atas nilai ambang cenderung kepada nilai '1', dan nilai di bawah nilai ambang cenderung kepada nilai '0'.

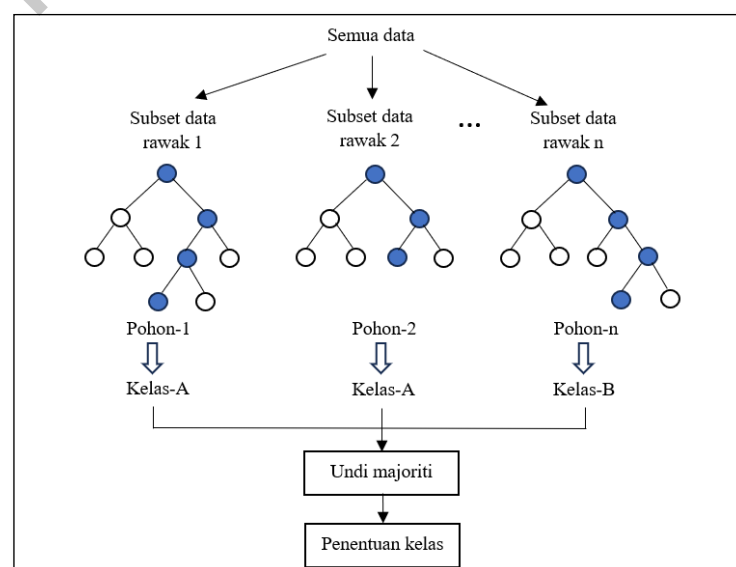


Rajah 2.9 Ilustrasi RL

f. Hutan Rawak

Hutan Rawak (HR) adalah algoritma pembelajaran berdasarkan konsep ensemble di mana ia merupakan gabungan daripada beberapa PK untuk menyelesaikan masalah yang kompleks dan untuk meningkatkan prestasi model.

Rajah 2.10 merupakan ilustrasi model pengelas HR. Setiap pokok keputusan menggunakan subset ciri dan subset data latihan yang dipilih secara rawak. Subset data rawak ini akan membolehkan setiap pokok menghasilkan pokok yang berbeza. Setiap pokok akan menghasilkan kelas sasaran. Kelas dengan undian terbanyak atau majoriti akan dipilih sebagai kelas sasaran yang akhir.



Rajah 2.10 Ilustrasi model pengelas HR