

PENGELASAN SOALAN PEPERIKSAAN BERLANDASKAN TAKSONOMI BLOOM MENGUNAKAN PEMBELAJARAN MESIN

Yeo Yong Sheng & Nazlia Omar*

**Pusat Kajian Teknologi Kecerdasan Buatan
Fakulti Teknologi & Sains Maklumat
43600 Universiti Kebangsaan Malaysia
(Corresponding author: nazlia@ukm.edu.my)**

Abstrak

Peperiksaan adalah perlu untuk menguji tahap kefahaman atas sesuatu perkara atau bidang. Untuk menjana soalan peperiksaan yang berkesan, pengajar haruslah mempunyai garis panduan untuk membina soalan yang seimbang daripada tahap kognitif yang berlainan yang dapat menilai pelajar dengan efektif. Garis panduan Taksonomi Bloom adalah antara garis panduan popular yang digunapakai oleh pengajar pada hari ini. Walau bagaimanapun, pengklasifikasian secara manual berdasarkan Taksonomi Bloom adalah satu perkara yang amat mencabar dan memerlukan masa yang panjang. Justeru, kajian ini mencadangkan satu model pengklasifikasian soalan peperiksaan berlandaskan taksonomi Bloom menggunakan teknik pembelajaran mesin. Pengelas yang digunakan dalam kajian ini ialah Mesin Vektor Sokongan (MVS), Bayes Naif (BN), Hutan Rawak (*Random Forest*), dan Jiran K-Terdekat (JKT). Untuk mendapatkan ketepatan yang lebih tinggi, set soalan data perlu dilakukan pra-pemprosesan seterusnya ciri pengekstrakan seperti beg perkataan digunakan. Satu laman sesawang yang mesra pengguna dibangunkan bagi memudahkan pengajar untuk mengklasifikasikan soalan peperiksaan mereka dengan mudah dan cepat serta melihat hasil analisis dari pengelas. Prototaip dari kajian ini dapat membantu para pengajar untuk menganalisis soalan peperiksaan bagi memenuhi keperluan untuk tahap kognitif yang berbeza bagi pelajar mengikut tahap pengajian.

Kata kunci: Pembelajaran mesin; pengelasan soalan; taksonomi bloom

Abstract

An examination is needed to evaluate how much the students understand what is being thought. In order to have a quality and effective examination, instructors should have a guideline on how to develop quality questions from different cognitive levels in a balanced way that can evaluate the students. Thus, many instructors today use Bloom's Taxonomy which is a framework that developed to assess students' intellectual abilities and skills. However, it is quite challenging for the examiner to classify out the question based on Bloom's Taxonomy manually. Hence, this study aims to propose a question classification model to classify question based on Bloom's Taxonomy cognitive domain by using machine learning approach. The classifiers used are Support Vector Machine (SVM), Naïve Bayes (NB), Random Forest (RF) and K-Nearest Neighbour (KNN). To obtain a more accurate result, the dataset collected will undergo pre-process text and feature extraction such as bag-of-words. A user-friendly website was developed to facilitate instructors to classify their exam questions easily and quickly including viewing the analysis results from the classifier. The prototype from this study can help instructors to analyze exam questions to meet the needs for different cognitive levels for students according to the level of study.

Keywords: Bloom's taxonomy; machine learning; question classification

1.0 PENGENALAN

Peperiksaan perlu dirangka dengan berkesan supaya pemantapan budaya inovasi melalui penilaian tahap keupayaan seseorang dalam memahami ilmu yang dipelajari. Kertas peperiksaan yang berkualiti haruslah terdiri daripada pelbagai aspek dan tahap kesukaran untuk menguji dan menilai tahap pemahaman pelajar dalam sesuatu perkara yang diajar. Oleh itu, soalan peperiksaan haruslah seimbang daripada setiap aspek yang bersesuaian dengan isi kandungan yang diajar dan memenuhi objektif pembelajaran. Taksonomi dalam bidang pendidikan telah memberi kesedaran kepada para pengajar sejak lahirnya kurikulum bahawa tujuan pembelajaran harus dianalisis sebelum proses pembelajaran dan pengajaran berlangsung bagi mengenalpasti silibus yang penting dan bersesuaian dengan tahap pelajar.

Benjamin Bloom (1956) iaitu seorang ahli psikologi dalam bidang pendidikan telah memperkenalkan satu hirarki kognitif yang dikenali sebagai taksonomi Bloom pada hari ini. Beliau telah mengklasifikasikan aras tingkah laku keintelektualan dalam pembelajaran kepada tiga domain iaitu domain kognitif, domain afektif, dan domain psikomotor dalam tujuan pendidikan ini. Taksonomi Bloom lebih menitikberatkan dalam aspek kognitif atau pengetahuan seperti fakta, teori dan hukum yang dapat membantu guru dalam merancang

pembelajaran dan pengajaran serta membuat soalan pelbagai aras dalam menguji tahap berfikir pelajar. Taksonomi Bloom bertujuan untuk membina pemikiran, pemahaman dan kebolehan secara seimbang. Taksonomi Bloom mempunyai enam tahap kognitif yang setiap tahap mewakili serta menilai kemahiran yang berbeza. Bagi menghasilkan peperiksaan yang lebih berkualiti, para pengajar telah dicadangkan untuk membangunkan soalan peperiksaan berdasarkan taksonomi Bloom (Ghanem et al. 2013).

Sehingga kini, Taksonomi Bloom masih lagi diguna pakai sebagai garis panduan yang penting dalam sektor pendidikan (Stringer et al. 2021; Curzon & Tummons, 2013). Walau bagaimanapun, terdapat juga segelintir soalan peperiksaan yang tidak dapat menggunakan garis panduan Taksonomi Bloom dengan tepat dan telah menyebabkan peperiksaan tersebut tidak dapat mencapai kesan yang terbaik kerana tidak seimbang daripada garis panduan Taksonomi Bloom (Haris, Sufi & Omar, 2015; Swart, 2010). Hal ini kerana 6 peringkat kognitif berbeza dalam Taksonomi Bloom mempunyai kata kunci mereka yang tersendiri. Kebanyakan kata kunci ini terdiri dari kata kerja. Kata kunci dalam setiap peringkat meliputi skop yang agak besar dan ini telah menyebabkan kesusahan untuk seseorang mengklasifikasikannya secara manual. Selain itu, alatan yang terdapat untuk mengklasifikasikan soalan memerlukan pakar komputer untuk mengendalikannya dan ianya amat terhad dan tiada di pasaran.

Tujuan utama kajian ini adalah untuk membangunkan satu sistem pengelasan soalan peperiksaan Bahasa Inggeris kepada 6 kognitif yang terdapat dalam Taksonomi Bloom. Untuk mencapai objektif utama kajian ini, langkah pertama yang dijalankan ialah membangunkan model pembelajaran mesin dengan data yang dikumpul menggunakan pemprosesan bahasa tabii untuk mengklasifikasikan soalan peperiksaan agar hasilnya boleh menepati tahap kognitif yang berbeza. Langkah kedua melibatkan pembangunan satu aplikasi web berintegrasi model pembelajaran mesin yang dipilih. Web aplikasi ini akan mengklasifikasikan dan menganalisis soalan peperiksaan yang disediakan.

2.0 PENGELASAN SOALAN BERDASARKAN TAKSONOMI BLOOM

Jayakodi et al. (2015) menggunakan satu set data yang mengandungi 88 soalan peperiksaan (53 data latihan dan 35 data ujian) yang diambil dari beberapa kursus teknologi maklumat. Kaedah WordNet dan persamaan kosinus telah digunakan untuk membangunkan algoritma pengelasan dengan kaedah berasaskan peraturan. Soalan peperiksaan yang tidak menggunakan kata kerja atau kata kunci dalam Taksonomi Bloom telah menyebabkan soalan tersebut dikategorikan ke tahap yang tidak tepat. Oleh itu, penggunaan persamaan kosinus atas corak soalan peperiksaan berasaskan peraturan telah membantu untuk

mengklasifikasikan soalan tersebut dengan betul. Mereka telah mencapai ketepatan sebanyak 70% melalui kaedah ini. Ketepatan masih boleh dipertingkatkan lagi dengan menambahbaik corak soalan peperiksaan supaya nilai dari persamaan kosinus akan lebih tepat (Jayakodi 2019).

Haris, Sufi & Omar (2015) telah mengumpul set data yang merangkumi 135 soalan peperiksaan (100 data latihan dan 35 data ujian) dan menggunakan kaedah berasaskan peraturan untuk mengklasifikasikan soalan peperiksaan. Kajian tersebut telah mencapai ketepatan 77% dan pengkaji telah mencadangkan untuk menetapkan aturan yang lebih banyak supaya dapat mencapai ketepatan yang lebih tinggi pada masa kelak (Haris & Omar 2015).

Yusof dan Jin (2010) telah menggunakan kaedah *Artificial Neural Network* (ANN) untuk mengklasifikasikan set data sebanyak 274 soalan peperiksaan (192 data latihan dan 82 data ujian). Kajian ini fokus pada *Document Frequency*(DF) dan *Category Frequency-Document Frequency*(CF-DF). Walau bagaimanapun, kaedah ini hanya mencapai ketepatan setinggi 65.9% sahaja.

Osadi et al. (2017) telah menggabungkan 4 kaedah pengelasan iaitu MVS, JKT, BN dan kaedah berasaskan peraturan untuk membina algoritma bagi pengelasan soalan peperiksaan. Set data yang dipakai merangkumi 60 data latihan dan 40 data ujian soalan peperiksaan. Pengkaji menggunakan pengekstrakan ciri berasaskan perwakilan vektor untuk kaedah berasaskan peraturan dan pengekstrakan ciri pelabelan golongan kata untuk kaedah pembelajaran mesin. Kaedah penggabungan ini telah mencapai ketepatan yang baik iaitu 82% dan pencapaian ini lebih tinggi berbanding dengan penggunaan pengelas individu.

Pincay & Ochoa (2015) menggunakan teknik perlombongan teks (*text mining*) dan BN untuk pengklasifikasian soalan peperiksaan. Dapatan kajian menunjukkan bahawa keputusan kajian bergantung kepada kualiti set latihan iaitu kekurangan set latihan akan mengakibatkan ramalan sistem yang kurang tepat. Keputusan kategori set data ujian dibandingkan dengan kategori set data ujian yang sebenarnya. Sebanyak 180 jawapan forum perbincangan dalam Bahasa Sepanyol telah digunakan. Nilai dapatan untuk tahap kognitif penilaian dan aplikasi dalam kajian mendapat 100%. Walau bagaimanapun, tahap kognitif lain mencapai nilai dapatan kurang daripada 60%. Purata ketepatan kajian adalah sangat rendah dengan nilai 20% sahaja.

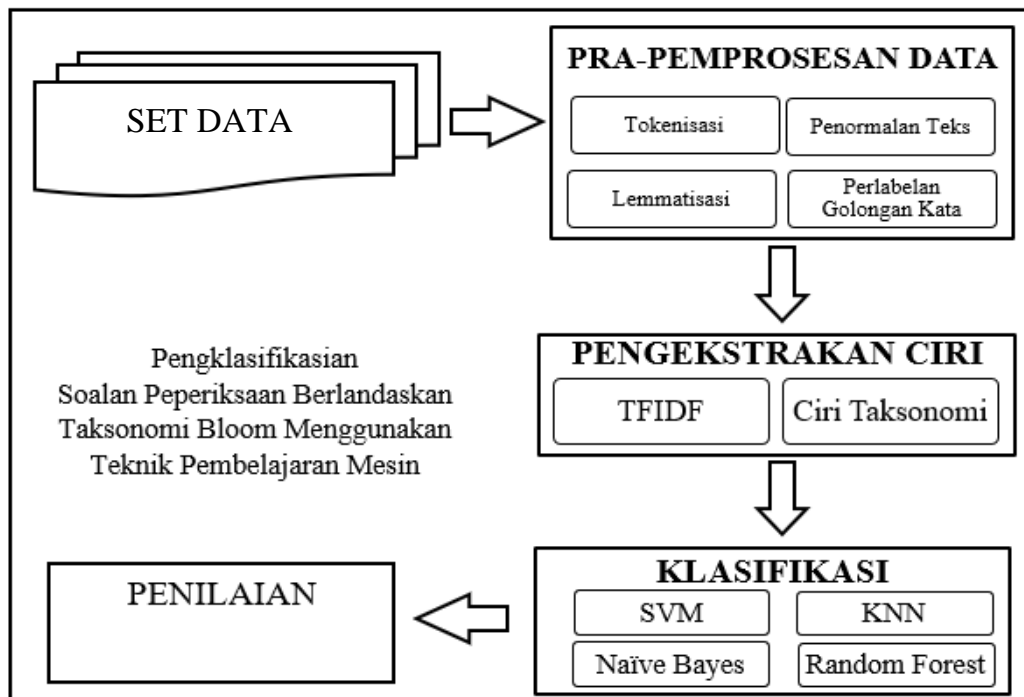
Jain et al. (2019) menggunakan pelbagai teknik pembelajaran mesin untuk mengklasifikasikan kertas soalan ke dalam tahap taksonomi Bloom yang berbeza. Set data yang digunakan terdiri daripada 1024 soalan yang diperolehi dari tiga universiti. Pengelas yang digunakan adalah Jiran K-Terdekat (JKT), Hutan Rawak, Pokok Keputusan (PK), Mesin Vektor Sokongan (MVS), Rangkaian Neural, Analisis Diskriminan Linear (ADL) dan Regresi Logistik untuk menilai pendekatan yang digunakan. Keputusan menunjukkan teknik Regresi Logistik dan Analisis Diskriminasi Linear (ADL) adalah yang terbaik dengan kedua-duanya mempunyai ketepatan 83.3%. Namun tiada ciri khusus yang digunakan semasa pengekstrakan data. Perlu kesimpulan berdasarkan komen ini: Topik 2.0 memerlukan kesimpulan dari hasil kajian literatur bagi memberi kefahaman kepada pembaca dan memberi justifikasi kenapa pengelas dan ciri.

Kesimpulannya, penggunaan pembelajaran mesin bagi pengelasan soalan berdasarkan Taksonomi Bloom merupakan kaedah yang popular dan terkini. Penggunaan Mesin Vektor Sokongan (MVS), Bayes Naif (BN), Hutan Rawak (*Random Forest*), dan Jiran K-Terdekat (JKT) dapat menghasilkan ketepatan yang tinggi. Kajian ini akan mengguna dan membandingkan empat kaedah mesin pembelajaran ini dengan kaedah pengekstrakan ciri yang tidak pernah dijalankan dalam kajian lepas.

3.0 METODOLOGI

Rajah 1 menunjukkan seni bina pengklasifikasian soalan peperiksaan berlandaskan Taksonomi Bloom menggunakan teknik pembelajaran mesin. Bagi merealisasikan pembangunan sistem ini, beberapa fasa kajian yang terlibat akan diterangkan dengan lebih lanjut.

Fasa Permulaan Kajian iaitu perancangan projek pada awal kajian yang telah dijalankan untuk mengenalpasti isu yang terlibat dan mengkaji teknologi terkini serta kajian lepas. Kajian ini seterusnya terdiri daripada empat fasa iaitu Fasa Pengumpulan Data, Fasa Pra-Pemrosesan Data, Fasa Pengekstrakan Ciri, Fasa Pembinaan Model Pengelasan, Fasa Pengujian dan akhirnya Fasa Pembangunan Web Aplikasi untuk mencapai objektif kajian.



Rajah 1. Seni bina sistem

3.1 Fasa Pengumpulan Data

Kajian ini menggunakan dua domain set data terbuka untuk memastikan kestabilan dan keberkesanan teknik pengelasan dalam kajian ini. Soalan set data yang dikumpulkan merangkumi soalan penuh tanpa pengekstrakan ciri-ciri tahap kognitif. Set data yang pertama terdiri dari 141 soalan dari tahap kognitif Taksonomi Bloom yang berbeza yang dikumpul dari laman web, buku dan juga kajian lepas (Haris & Omar, 2015). Set data kedua merangkumi 600 soalan (Yahya et al. 2012) yang merupakan data standard yang digunakan oleh kajian lain. Kedua-dua set data ini telah dahulu dilabelkan kepada enam tahap kognitif dalam Taksonomi Bloom iaitu: (1) *knowledge* (pengetahuan), (2) *comprehension* (pemahaman), (3) *application* (penerapan), (4) *analysis*(analisis), (5) *synthesis*(sintesis) dan (6) *evaluation*(penilaian). Jadual 1 menunjukkan contoh soalan dalam set data.

Jadual 1. Contoh soalan setiap tahap kognitif Taksonomi Bloom dalam set data

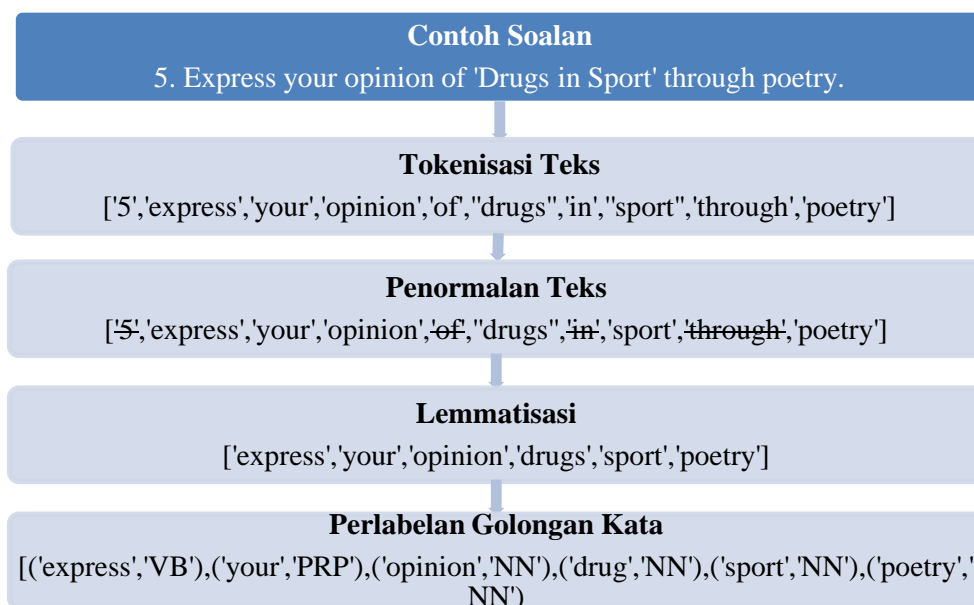
Tahap Kognitif	Contoh Soalan Bahasa Inggeris dalam set data
Knowledge (Pengetahuan)	Define Inheritance concept. Name five cities in US. State how to find a word in a dictionary.
Comprehension (Pemahaman)	Identify three mistakes from the passage and correct them. Describe how interest rates affect the economy. Explain in your own words what a recessive gene is.
Application (Penerapan)	Sketch a diagram which shows these fractions or take photographs of the fractions. Interpret the graph and state how many trees were cut down to produce paper. Draw a picture of the gown Cinderella wore to the ball.
Analysis (Analisis)	Analyze the following questionnaire. What is the relationship between probability and statistical analysis? Compare and contrast the waterfall model with the prototyping model.
Synthesis (Sintesis)	Rewrite the story from an animal's point of view. Write a journal from the point of view of mountaineer. Create and perform a play about frogs.
Evaluation (Penilaian)	Judge how well a project meets the criteria of a rubric. Judge the validity of arguments for and against astrology. Decide whether you learned enough about electricity from this book.

3.2 Fasa Pra-Pemprosesan Data

Pra-pemprosesan data bagi data yang tidak berstruktur amat penting sebelum soalan diproses menggunakan kaedah pemprosesan bahasa tabii bagi menyingkirkan data yang tidak penting, berganda atau tidak relevan supaya dapat memudahkan pengendalian data untuk proses seterusnya.

Tokenisasi ialah satu proses digunakan untuk memecahkan data ayat menjadi data perkataan. Kajian ini telah mendapati bahawa pakej *Keras* dalam Python boleh menghasilkan tokenisasi yang paling baik kerana ia berupaya memberikan set data yang telah ditukar menjadi huruf kecil dan pemansuhan tanda baca menjadikan hasil bilangan token yang lebih kecil. Ini dapat memudahkan lagi proses pra-pemprosesan data seterusnya. Pemansuhan nombor dan kata henti (contohnya perkataan *the*) akan dilakukan dalam penormalan teks. Walau bagaimanapun, bukannya semua kata henti akan digunapakai dalam kajian ini kerana sesetengah kata henti dalam soalan seperti *what*, *when*, *where*, *which*, *how* dan lain-lain memainkan peranan penting dalam klasifikasi tahap kognitif.

Seterusnya proses lemmatisasi dijalankan yang merupakan satu proses dalam menyingkirkan kata imbuhan awalan ataupun kata imbuhan akhiran dalam satu perkataan dan mendapatkan lemma perkataan. Langkah terakhir dalam pra-pemprosesan data ialah Perlabelan Golongan Kata (PGK) yang merupakan proses pemberian penanda atau label pada golongan kata setiap perkataan di dalam korpus. PGK dapat membantu dalam proses pengekstrakan ciri seterusnya. Rajah 2 menunjukkan set data selepas pra-pemprosesan data secara lengkap.



Rajah 2. Contoh data diproses dalam fasa Pra-pemprosesan data

3.3 Fasa Pengekstrakan Ciri

Dalam kaedah pembelajaran mesin, setiap set data pada mulanya hendaklah diubah menjadi perwakilan vektor dahulu untuk mengklasifikasikannya (Marr, 2017). Proses ini dipanggil pengekstrakan ciri. *Bag-of-Words* ialah antara cara yang umum untuk memproses bahasa semula jadi, yang setiap ciri ialah satu perkataan dalam korpus, yang boleh bertukar menjadi nilai Boolean, menunjukkan sama ada perkataan tertentu wujud dalam soalan, dan juga boleh menjadi integer atau nombor nyata, serta menunjukkan frekuensi setiap perkataan tanpa mengira posisi perkataan tersebut. Setiap perkataan dalam ayat akan diekstrakan menjadi ciri dan dokumen tersebut akan bertukar menjadi perwakilan dalam bentuk vektor.

Teknik yang ditambahbaik daripada *Bag-of-Words* yang digunakan untuk pengekstrakan ciri data teks adalah TF-IDF (*term frequency-inverse document frequency*). TF-IDF merupakan ukuran statistik yang menggambarkan pentingnya suatu istilah terhadap sebuah dokumen dalam sebuah kumpulan atau korpus. Ciri TF-IDF telah digunakan dalam kajian ini.

Kata kerja dalam setiap soalan merupakan kata kunci dalam proses pengelasan kerana ia memainkan peranan yang sangat penting dalam klasifikasi soalan tersebut ke tahap kognitif taksonomi Bloom. Kata kunci biasanya akan terdapat pada permulaan setiap soalan ataupun di tengah-tengah setiap soalan. Kadangkala satu soalan akan mempunyai lebih daripada satu kata kunci. Oleh itu, ciri taksonomi ialah mendapatkan dan menggunakan setiap kata kunci yang terdapat dalam soalan tersebut seperti yang ditunjukkan dalam Jadual 2.

Jadual 2. Ciri Taksonomi

Contoh Soalan	Interpret and discuss the concept that can be explained from the graph theory.
Ciri Taksonomi:	{interpret, discuss}
Bag of Words Selepas Ciri Taksonomi:	{interpret, discuss, concept, explain, graph, theory, interpret}

3.4 Fasa Pembinaan Model Pengelasan

Proses yang dijalankan pada fasa ini adalah membangunkan model pengelas yang sesuai bagi menyelesaikan masalah yang dikenalpasti dalam kajian ini. Untuk mengklasifikasikan soalan peperiksaan berlandaskan Taksonomi Bloom, kaedah pembelajaran mesin diselia yang telah digunakan iaitu MVS, JKT, BN dan Hutan Rawak digunakan dalam fasa ini. Parameter lalai telah digunakan untuk setiap kaedah. Set data yang sedia ada dijadikan 80% data latihan dan 20% data ujian. Ketepatan empat model ini akan dibandingkan dan kajian ini akan menggunakan model yang boleh mendapatkan ketepatan yang tertinggi. Fasa ini dibahagikan kepada dua langkah utama iaitu melatih model pada mulanya dan kemudian menguji model tersebut. Kedua-dua langkah ini akan menggunakan set data dengan mengagihkan set data yang sedia ada menjadikan 80% data latihan dan 20% data ujian.

3.5 Fasa Pengujian

Fasa pengujian ini akan menilai prestasi model pengelasan menggunakan parameter yang berlainan seperti kejituan (*precision*), dapatan (*recall*) dan nilai-F. Beberapa istilah harus dikenali sebelum menggunakan parameter tersebut dalam kajian ini iaitu Positif Benar (TP), Positif Palsu (FP) dan Negatif Palsu (FN). Positif Benar (TP) ialah nombor soalan yang diklasifikasikan dengan betul oleh sistem berdasarkan tahap kognitif Taksonomi Bloom, Positif Palsu (FP) ialah nombor soalan yang diklasifikasikan dengan salah manakala Negatif Palsu (FN) ialah nombor soalan yang tidak diklasifikasikan. Nilai pengujian yang didapatkan dari parameter kejituan (*precision*), dapatan (*recall*) dan ukuran-F adalah di antara 0 dan 1. Semakin tinggi nilai pengujian yang semakin dekat dengan nilai 1 bermaksud keberkesanan sistem yang semakin baik dan sebaliknya semakin rendahnya nilai pengujian yang semakin dekat dengan nilai 0 bermaksud keberkesanan sistem yang lemah.

3.6 Fasa Pembangunan Aplikasi Web

Selepas memperoleh model pembelajaran mesin yang terbaik, langkah seterusnya adalah membina aplikasi web yang dapat menghubungkan model dan menghasilkan ramalan mengenai data baru dalam masa sebenar. Oleh itu, kajian ini telah menggunakan "Gradio" untuk membangunkan web aplikasi. Gradio ialah antara pustaka Python sumber terbuka untuk mencipta komponen antara muka yang mudah digunakan. Ia boleh disesuaikan untuk model pembelajaran mesin, sebarang API (*application programming interface*) dengan pelbagai fungsi. Dengan adanya web aplikasi, pengguna atau pengajar dapat memuat naik soalan peperiksaan dan melihat hasil analisis soalan tersebut mengikut pengelasan tahap kognitif taksonomi Bloom.

4.0 HASIL KAJIAN

Berikut merupakan prestasi model-model pengelasan seperti perbincangan di atas dan hasil ketepatan dibandingkan seperti dalam Jadual 3.

Jadual 3. Ketepatan model pengelasan

Model Pengelasan	Mesin Vektor Sokongan (MVS)	Jiran K-Terdekat (JKT)	Bayes Naif (BN)	Hutan Rawak
Ketepatan	77.18%	76.51%	70.47%	72.48%

MVS adalah salah satu algoritma pembelajaran mesin yang digunakan dalam klasifikasi dan regresi. Algoritma ini bekerja dengan membangun sebuah model yang dapat memprediksi kelas atau nilai dari data yang diberikan. MVS mencari sebuah *hyperplane* atau bidang pemisah yang dapat memisahkan data menjadi dua kelas dengan *margin* (jarak) maksimum antara kelas tersebut.

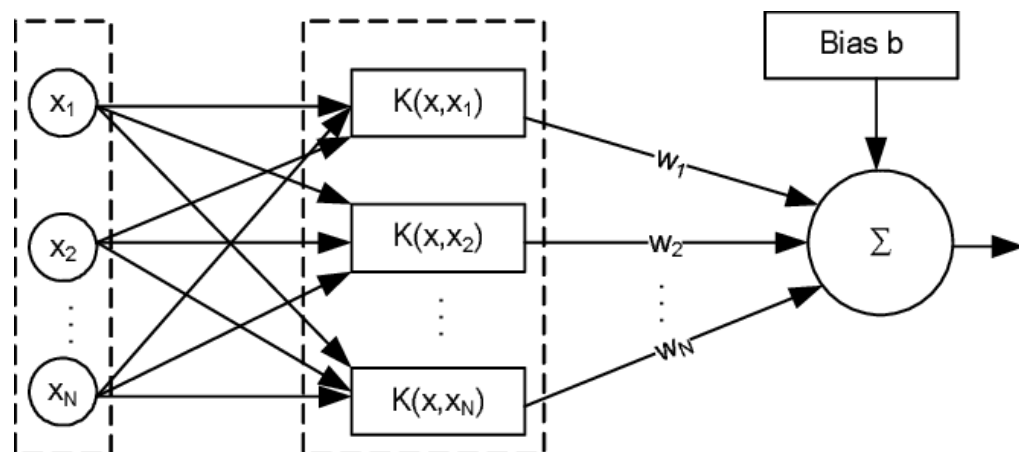
MVS dapat memisahkan kelas yang sangat kompleks dan tidak teratur, bahkan ketika data tidak sepenuhnya linier terpisah. MVS menggunakan fungsi kernel untuk memetakan data ke dalam ruang dimensi yang lebih tinggi, di mana data dapat terpisah dengan lebih baik. MVS dapat menangani data yang tidak seimbang, di mana jumlah sampel dalam kelas yang berbeda tidak sama. MVS mengoptimalkan pemisahan kelas dengan mengidentifikasi batas keputusan yang tepat, yang meminimalkan jumlah sampel yang salah dikelompokkan.

Kelemahan JKT adalah ia sangat bergantung pada jumlah dan kualiti data latihan. Jika data latihan yang digunakan tidak wakil atau tidak mencakupi variasi yang mencukupi, maka JKT akan menghasilkan ramalan yang tidak tepat. Selain itu, JKT juga boleh menjadi lambat dalam menghasilkan ramalan apabila data latihan yang digunakan sangat besar dan rumit. JKT juga boleh terpengaruh oleh nilai pencilan dalam data latihan.

Di samping itu, BN menganggap bahawa setiap atribut (fitur) saling bebas atau tidak berkaitan satu sama lain adalah kelemahannya. Oleh itu, jika atribut dalam data latihan berkaitan atau saling mempengaruhi, maka hasil prediksi yang dihasilkan oleh algoritma ini boleh menjadi tidak tepat. Selain itu, BN juga boleh terpengaruh oleh atribut yang tidak relevan dalam data latihan dan kebolehgunaan model dapat menurun apabila terdapat banyak atribut yang tidak relevan.

Kelemahan dari Hutan Rawak adalah ia cenderung *overfitting* apabila terlalu banyak pohon keputusan yang dibina, terutamanya apabila data latihan yang digunakan sedikit dan jumlah fitur yang digunakan terlalu banyak. Overfitting berlaku apabila model terlalu menyesuaikan diri dengan data latihan dan kebolegunaan model untuk melakukan ramalan pada data baru menurun. Selain itu, Hutan Rawak juga boleh mengambil masa yang lebih lama untuk melatih model berbanding dengan beberapa algoritma pembelajaran mesin yang lain.

Berdasarkan Jadual 3, model pengelasan MVS menunjukkan ketepatan yang paling tinggi di antara model pengelasan yang digunakan seperti yang dinyatakan MVS dapat memisahkan kelas yang sangat kompleks dan tidak teratur, bahkan ketika data tidak sepenuhnya linier terpisah. Rajah 3 telah menunjukkan seni bina model MVS yang terpilih sebagai model pengelasan terbaik yang akan digunakan untuk pembinaan web aplikasi.

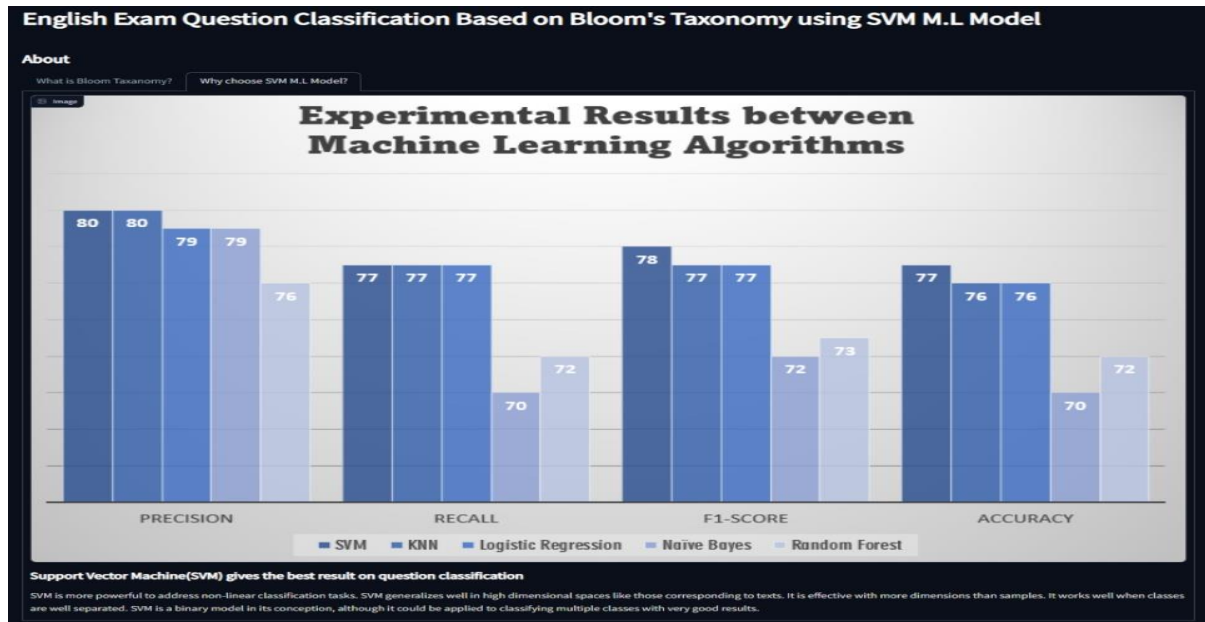


Rajah 3. Seni bina model MVS

Ciri taksonomi dan ciri TF-IDF sebagai pengekstrakan ciri dalam pemrosesan data dapat meningkatkan lagi keberkesanan pengelasan model. Keputusan ini tidak dapat dibandingkan dengan kajian lain yang berkaitan kerana data yang digunakan adalah berlainan. Berdasarkan keputusan yang diperolehi, sistem ini seterusnya dibina menggunakan model pengelasan MVS.

Web aplikasi yang dibina dibahagikan kepada dua bahagian iaitu bahagian informasi dan bahagian pengelasan soalan. Taksonomi Bloom diterangkan pada bahagian informasi supaya pengguna dapat mengenali panduan untuk membina soalan yang dapat memenuhi tahap kognitif Taksonomi Bloom. Kaedah pembelajaran mesin yang digunakan untuk

membina pengklasifikasian soalan juga diterangkan secara umum dan membandingkan kaedah pembelajaran mesin yang telah diuji sebelum membangunkan web aplikasi seperti yang ditunjukkan dalam Rajah 4.



Rajah 4. Keputusan Model Pembelajaran Mesin pada web aplikasi

Untuk mengenalpasti tahap kognitif Taksonomi Bloom soalan peperiksaan, pengguna boleh menaip soalan peperiksaan satu demi satu dahulu sebelum diletakkan dalam set soalan peperiksaan seperti yang ditunjukkan dalam Rajah 5.

Start classifying your exam question/s now~

Question

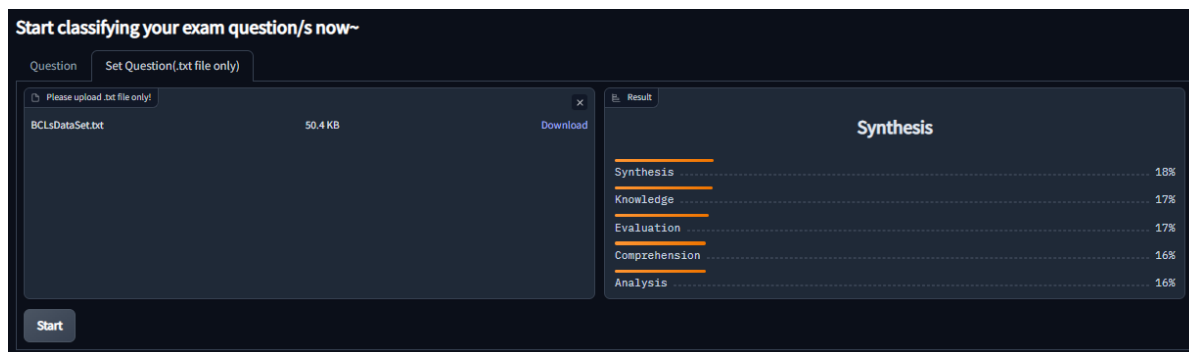
Please type your question here:

Result

Rajah 5. Hasil pengklasifikasian satu soalan peperiksaan

Pengguna juga dapat mengklasifikasikan set soalan peperiksaan yang disediakan dalam format *.txt* dan dapat mengetahui pembahagian tahap kognitif untuk soalan peperiksaan yang disediakan seperti Rajah 6. Jumlah soalan tidak dihadkan bagi set soalan ini dan boleh terdiri dari pelbagai tahap kognitif Bloom. Aplikasi ini akan menganalisis setiap

soalan di dalam set tersebut dan memaparkan hasil tahap kognitif bagi kesemua soalan tersebut di dalam bentuk peratusan. Rajah 6 telah menggunakan sebahagian dari dataset yang diguna dalam kajian ini untuk menguji hasil web aplikasi yang dibina. Hasil analisis ini amat berguna bagi pengajar dalam membantu menyediakan soalan peperiksaan yang bersesuaian mengikut tahap pengajian pelajar.



Rajah 6. Hasil pengklasifikasian satu set soalan peperiksaan

5.0 KESIMPULAN

Kajian ini menunjukkan bahawa penggunaan kaedah pembelajaran mesin dapat memberikan ketepatan yang baik bagi pengelasan soalan berdasarkan tahap kognitif taksonomi Bloom. Pengekstrakan ciri seperti ciri taksonomi dan TF-IDF dapat menambahbaik prestasi pengklasifikasian sistem. Selain itu, web aplikasi ini sangat mudah digunakan oleh semua pengguna kerana pengguna boleh memilih untuk mengklasifikasikan set soalan peperiksaan ataupun secara memasukkan satu demi satu soalan peperiksaan. Namun, kekangan sistem ini ialah ianya hanya terhad untuk menganalisis soalan peperiksaan dalam Bahasa Inggeris dan juga set soalan hendaklah dalam bentuk .txt format sebelum dimuat naik ke sistem untuk dianalisis. Untuk penambahbaikan masa hadapan sistem ini, boleh ditambahbaik lagi dengan menambah kepelbagaian bahasa soalan peperiksaan untuk dianalisis seperti Bahasa Melayu. Sistem ini dibangunkan bertujuan untuk memanfaatkan pengajar agar dapat mengklasifikasi dan menganalisis soalan peperiksaan yang disediakan sesuai dengan tahap pengajian tersebut.

6.0 PENGHARGAAN

Kajian ini dibiayai oleh projek UKM PDI-2021-028

7.0 RUJUKAN

Bloom, B. (1956). *Taxonomy Of educational objectives: The Classification of educational goals*. London: Longman.

Curzon, L. and Tummons, J. (2013). *Teaching in further education: An Outline of principles and practice*. Seventh edition. London: Bloomsbury.

Ghanem Nayef, E., Rosila, N., Yaacob, N. & Ismail, H. N. (2013). Taxonomies of educational objective domain. *International Journal of Academic Research in Business and Social Sciences*, 3(9), 2222–6990.

Haris, S.Sufi and Omar Nazlia (2015). A Rule-based approach in bloom's taxonomy question classification through natural language processing, *IEEE 2012*, pp. 410-414.

Haris, S. S. & Omar, N. (2015). Bloom's taxonomy question categorization using rules and n-gram approach. *Journal of Theoretical and Applied Information Technology* 76(3), 401–407.

Hassan, M. R., Hossain, M. M., Bailey, J., dan Ramamohanarao, K. (2008). Improving K-Nearest neighbor classification with distance functions based on receiver operating characteristics. *Lecture Notes in Computer Science Book Series (LNAI, Volume 5211)*.

Jayakodi, M. Bandara and I. Perera, (2015) An automatic classifier for exam questions in Engineering: A process for Bloom's taxonomy, *2015 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*, Zhuhai, China, 2015, pp. 195-202, doi: 10.1109/TALE.2015.7386043.

Jain, M., Beniwal, R., Ghosh, A., Grover, T., Tyagi, U. (2019). Classifying question papers with bloom's taxonomy using machine learning techniques. In: Singh, M., Gupta, P., Tyagi, V., Flusser, J., Ören, T., Kashyap, R. (eds) *Advances in computing and data sciences*.

ICACDS 2019. *Communications in Computer and Information Science*, vol 1046. Springer, Singapore.

Jayalakshmi, S. & Ananthi, (2015). Question classification: A Review of state of the art algorithms and approaches. *Indian Journal of Science and Technology*, pp. 1-4 8(29).

Jayakodi K., Bandara M., Perera. I, and Meedeniya D. (2016). WordNet and cosine similarity based classifier of exam questions using Bloom's taxonomy. *International Journal of Emerging Technologies in Learning (IJET)*, 11(04), pp. 142–149.

Osadi, K., Fernando, M. and Welgama, W. (2017). Ensemble classifier based approach for classification of examination questions into bloom's taxonomy cognitive levels. *International Journal of Computer Applications* 162(4),1-6.

Marr, B. (2017). Supervised V Unsupervised Machine Learning -- What's the difference? *Forbes*. Retrieved March 26, 2022, from <https://www.forbes.com/sites/bernardmarr/2017/03/16/supervised-v-unsupervised-machine-learning-whats-the-difference/?sh=532d2a96485d>.

Pincay, J., Ochoa, X. (2013). Automatic classification of answers to discussion forums according to the cognitive domain of blooms taxonomy using text mining and a bayesian classifier. In: *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications*, pp. 626–634.

Stringer, J. K., Santen, S. A., Lee, E., Rawls, M., Bailey, J., Richards, A., Biskobing, D. (2021). Examining Bloom's Taxonomy in multiple choice questions: students' approach to questions. *Medical Science Educator*, 31(4), 1311–1317.

Swart, A. J. (2010). Evaluation of final examination papers in engineering : A Case study using Bloom ' s Taxonomy. 53(2), pp. 257-264, *IEEE Transactions on Education*.

Yahya, A. A., Toukal, Z. & Osman, A. (2012). Bloom's Taxonomy – Based classification for item bank questions using support vector machines bloom ' s taxonomy – Based classification for item bank. *Modern Advances in Intelligent Systems and Tools, Studies in Computational Intelligence* 431, pp. 135–140.

Yusof N. and Jing H.C(2010). Determination of blooms cognitive level of questions item using artificial neural network. *International Conference on Intelligent System Design and Applicatons*, pp. 866-870.