

**PENGELASAN BERKAITAN KONSEP AL-TARGHIB (GALAKAN)
DAN AL-TARHIB (AMARAN) BERDASARKAN TERJEMAHAN AL-
QURAN DALAM BAHASA INGGERIS MENGGUNAKAN
PEMBELAJARAN MESIN**

SITI NURIZZATI BINTI AZMAN
DR. SAIDAH BINTI SAAD

^{1,2}*Fakulti Teknologi & Sains Maklumat, Universiti Kebangsaan Malaysia, 43600 UKM
Bangi, Selangor Darul Ehsan, Malaysia*

Abstrak

Al-Quran merupakan sumber utama pengetahuan dan panduan bagi semua umat Islam. Kaedah pengelasan teks cuba diaplikasikan untuk menyelesaikan masalah mengenai konsep Al-Targhib (galakan) dan Al-Tarhib (amaran) untuk menjadikan setiap tafsir di dalam Al-Quran dapat difahami dengan lebih baik oleh pembaca. Oleh itu, matlamat kajian ini adalah untuk membina model yang boleh mengklasifikasikan ayat-ayat Al-Quran dalam terjemahan Bahasa Ingggris kepada galakan dan amaran menggunakan teknik pembelajaran mesin iaitu pengelasan binari. Pendekatan pembelajaran mesin seperti pendekatan linguistik Mesin Sokongan Vektor (SVM), *Naïve Bayes* (NB) dan *K-Nearest Neighbour* (KNN), *Logistic Regression* (LR), *Decision Tree* (DT) dan *Random Forest* (RF) digunakan untuk mencapai matlamat ini, di samping kesan penggunaan ciri-ciri yang berbeza seperti *Term Frequency-Inverse Document Frequency* (TF-IDF), *Word2Vec*, dan *Bag of Words* (BOW) dalam pemendaman perkataan juga dinilai. Korpus beranotasi daripada Al-Quran juga dibina. Ia terdiri daripada 2000 ayat, iaitu 1000 ayat galakan dan 1000 ayat amaran. Keputusan dan analisis yang dijalankan menunjukkan ciri gabungan TF-IDF + BOW dan model pengelas LR mendapatkan prestasi yang terbaik. Gabungan kedua-dua ciri dan model terbaik ini memperoleh ketepatan pengelasan yang tertinggi iaitu sebanyak **86%**. Kajian ini merupakan kajian baru untuk mengekstrak maklumat dan pengetahuan dari Al-Quran. Ini penting bagi kedua-dua penyelidik dalam kajian Islam serta penyelidik tidak khusus.

Kata kunci: Al-Quran, Pengelasan binari, Korpus beranotasi, Pembelajaran mesin, Pemilihan ciri

Pengenalan

Al-Quran ialah kalam Allah SWT yang telah diturunkan kepada Nabi Muhammad SAW melalui perantaraan Jibril a.s. untuk menjadi panduan umat manusia. Al-Quran merupakan salah satu dari 5 buah kitab yang diturunkan oleh Allah SWT. Al-Quran bukanlah buku akan tetapi mengandungi ilmu pengetahuan yang banyak dan pelbagai yang termaktub dalam ayat-ayat suci Al-Quran. Susunan Al-Quran terdiri daripada cantuman 28 huruf Hijaiyah di mana keseluruhan Al-Quran mengandungi 323,671 huruf, 77,437 kalimah, 6236 ayat, 114 surah 30 juzuk, 60 hizib dan 240 maqra'. Al-Quran mempunyai gaya maklumat yang unik dan berkias, maka pencarian dan capaian maklumat Al-Quran memerlukan perhatian khusus supaya maklumat yang diperolehi adalah betul dan bermakna. Di antara isi kandungan Al-Quran ini adalah mengandungi ayat-ayat yang memberikan ransangan agar mendorong ke arah ketaatan kepada Allah; dan amaran supaya kita menghindari diri daripada berbuat kejahatan. Berdasarkan kefahaman ini dan berdasarkan perkembangan teknologi semasa, satu pendekatan cuba diaplikasikan bagi membantu mengklasifikasikan Al-Quran kepada dua konsep utama iaitu galakan dan amaran.

Pengelasan adalah merujuk kepada proses mengklasifikasikan dokumen teks yang tidak berlabel ke dalam kategori yang telah ditetapkan berdasarkan kandungannya. (Adeleke et al., 2017). Beberapa contoh algoritma pengkategorian teks telah dibangunkan, antaranya seperti *Naïve Bayes*, Mesin Sokongan Vektor dan *K-Nearest Neighbor Logistic Regression*, *Decision Tree* dan *Random Forest*. Pengkategorian teks telah digunakan secara meluas dalam set data berasaskan teks pada hari ini. Dalam kajian ini, pengelasan korpus Al-Quran boleh meningkatkan kebolehcapaian korpus Al-Quran berdasarkan konsep Al-Targhib (galakan) dan

Al-Tarhib (amaran). Kajian ini memfokuskan kepada pengaplikasian algoritma pembelajaran mesin untuk menentukan ciri dan pengelasan yang terbaik untuk pengelasan korpus Al-Quran yang disediakan.

1) Objektif kajian

Kajian ini dilaksanakan adalah untuk membina model yang mampu mengelaskan teks sama ada sebagai galakan atau amaran untuk setiap ayat. Objektif kajian ini adalah untuk:

1. Menentukan ciri pengelasan terbaik dalam pembelajaran mesin.
2. Mengenalpasti model pengelasan terbaik untuk mengelaskan konsep galakan dan amaran.
3. Membangunkan sistem web yang membolehkan pengguna membuat pengelasan ayat-ayat Al-Quran dalam terjemahan Bahasa Inggeris

2) Skop

Dalam kajian ini, korpus khusus Al-Quran telah dibina di dalam Bahasa Arab. Ayat-ayat galakan dan amaran ini telah diterjemahkan kepada Bahasa Inggeris bagi memudahkan proses pengelasan kepada salah satu konsep Al-Tarhib (galakan) dan Al-Tarhib (amaran). Kajian ini memberi tumpuan kepada pengelasan ayat-ayat tersebut menggunakan pendekatan model pembelajaran mesin yang berbeza dan pada ciri-ciri yang berbeza. Seterusnya mengenalpasti ciri dan model pengelasan terbaik yang mampu memberikan ketepatan analisis yang tinggi. Pengelasan ini hanya memfokuskan kepada ayat yang mempunyai satu kelas sahaja pada satu ayat sama ada ayat amaran atau ayat galakan. Jika ayat mempunyai kedua-dua jenis kelas tersebut maka ayat ini tidak termasuk dalam proses pengelasan ini.

3) Justifikasi dan Kepentingan Kajian

Dalam pembelajaran mesin, pengelasan teks ialah algoritma pembelajaran diselia yang memainkan peranan penting untuk mengkategorikan pemerhatian kepada beberapa kelas utama. Matlamat pengelasan dalam teks adalah untuk mengklasifikasikan urutan teks ke dalam satu daripada beberapa kelas. Dalam kajian ini, pengelas binari yang sesuai akan digunakan untuk mengelaskan korpus Al-Quran kepada dua bahagian iaitu galakan dan amaran serta ketepatannya akan direkodkan. Sebelum melatih data, data akan diproses terlebih dahulu untuk mengekstrak ciri-ciri utama di dalamnya. Pengelas yang tidak sesuai akan menyukarkan dapatan kajian ketika melakukan pengelasan Al-Quran. Keadaan ini membuatkan pengguna keliru untuk memilih pengelas yang terbaik dalam melakukan pengelasan teks. Dalam kajian ini, model pengelas dan ciri yang terbaik akan dikenalpasti untuk pengelasan setiap ayat Al-Quran kepada konsep konsep utama iaitu Al-Tarhib (galakan) dan Al-Tarhib (amaran).

4) Sorotan Susastera

Terdapat beberapa contoh pengelasan dalam pembelajaran mesin sedia menjadi rujukan bagi perbandingan pengelas dan ciri dalam kajian ini. Kajian lepas dijadikan rujukan bagi mengenalpasti pengelas dan ciri yang terbaik untuk digunakan dalam pengelasan teks. Jadual 1 menunjukkan perbandingan bagi kajian lepas yang merujuk kepada kaedah pendekatan yang digunakan bagi domain ayat Al-Quran (Islam).

Rujukan	Kaedah Pendekatan	Domain	Bahasa	Model Pengelas	Pemilihan Ciri
AlHasani, H., Saad, S., & Kassim, J. (2018)	Pembelajaran Mesin menggunakan Penandaan POS, N-Gram dan Pemilihan Ciri.	Pengelasan Tarhib dan Tarhib menggunakan Analisis Sentimen pada Al-Quran.	Arab Klasik dan Inggeris	SVM dan NB	Penandaan POS, N-Gram dan Pemilihan Ciri berdasarkan korelasi.
A. Adeleke, N. Samsudin, A. Mustapha, S.	Pembelajaran Mesin menggunakan	Pelabelan Ayat Al-Quran (Syahadah	Inggeris	SVM, NB, KNN, Pokok Penentu(J48)	Perolehan Maklumat (IG), ChiSquare (CH)

Ahmad Khalid (2017)	Pengelas Teks Automatik (ATC)	dan Sembahyang)			
A. Adeleke, N. Samsudin (2018)	Teknik Pemilihan Ciri Hibrid untuk pengelasan ayat Al-Quran	Pengelasan ayat Al-Quran (Iman, Ibadah dan Akhlak)	Inggeris	NB, libSVM, KNN, Pokok Penentu (J48)	Perolehan Maklumat (IG), Pemilihan ciri berasaskan korelasi (CFS), Hibrid IG-CFS, TF-IDF berwajaran
Ulumudin, A Adiwijaya, M S Mubarak (2019)	Pengelasan berbilang label ayat Al-Quran	Pengelasan topik Al-Quran	Melayu	KNN	TF-IDF berwajaran
Geehan Sabah Hassan, Siti Khaotijah Mohammad, Faris Mahdi Alwan (2015)	Pengelasan tafsir Al-Quran menggunakan pembelajaran mesin	Pengelasan tafsir Al-Quran	Melayu	KNN	TF-IDF
Ahmed Alsaffar, Nazlia Omar (2014)	Klasifikasi Sentimen Melayu menggunakan pembelajaran mesin	Analisis Sentimen dalam Bahasa Melayu	Melayu	SVM, NB, KNN	Perolehan Maklumat (IG), Analisis Komponen Utama (PCA), <i>Relief-f</i> , Indeks Gini, <i>Chi-squared</i> , Ketidakpastian, Asas SVM.

Jadual 1 Ringkasan Perbandingan Kajian Lepas

AlHasani et al. (2018) menjalankan kajian tentang Al-Quran, sumber utama ilmu dan panduan bagi umat Islam. Al-Quran menggunakan mempunyai konsep Targhib (Galakan) dan Tarhib (Amaran) untuk menggesa manusia ke arah kebaikan dan memperingatkan mereka daripada kejahatan. Untuk mengklasifikasikan ayat Al-Quran sebagai galakan, amaran, atau neutral, para penyelidik menggunakan analisis sentimen dan teknik NLP. Mereka membina model menggunakan pendekatan pembelajaran mesin, termasuk penandaan POS, analisis N-Gram dan pemilihan ciri dengan kaedah berasaskan korelasi. Kajian itu melibatkan korpus beranotasi sebanyak 2000 ayat, yang terdiri daripada 1000 ayat galakan dan 1000 ayat amaran.

Naïve Bayes mencapai ketepatan 95.6% dan Mesin Vektor Sokongan mencapai ketepatan 91.5% menggunakan ciri 5000 teratas yang dipilih.

A. Adeleke, et. al (2017) telah membuat penyelidikan mengenai Penglabelan Ayat Al-Quran dalam Pembelajaran Mesin menggunakan Pengelas Teks Automatik (ATC). Tujuan kajian ini dijalankan adalah untuk membangunkan model yang boleh dilatih secara automatik untuk mengklasifikasikan ayat-ayat Al-Quran terpilih kepada tiga label kelas yang ditetapkan; iman, ibadah dan adab(akhlak). Empat pengelas ML konvensional telah digunakan iaitu SVM, NB, pokok keputusan (J48) dan KNN dengan menggunakan ciri Perolehan Maklumat (IG), ChiSquare (CH). Di akhir kajian, didapati pengelas mencapai skor ketepatan melebihi 80% dengan algoritma NB merekodkan skor tertinggi keseluruhan 93.9% ketepatan dan 0.964 luas kawasan di bawah lengkung(AUC).

A. Adeleke & N. Samsudin (2018) mengkaji tentang ketepatan Pengelasan Ayat-Al-Quran menggunakan Teknik Pemilihan Ciri Hibrid Perolehan Maklumat (IG) dan Pemilihan ciri berasaskan korelasi (CFS) iaitu IG-CFS. Pemilihan ciri sedia ada yang digunakan kini mempunyai masa jalan pengiraan yang tinggi dan mempunyai kejituan yang rendah. Oleh itu, teknik ciri hibrid IG-CFS digunakan untuk mencapai prestasi kejituan yang tinggi dan masa jalan pengiraan yang lebih singkat. Teknik IG-CFS digunakan untuk melabelkan ayat Al-Quran Al-Baqara dan Al-Anaam daripada terjemahan Bahasa Inggeris dan ulasan (tafsir). "StringToWordVector" dengan kaedah TF-IDF berwajaran telah digunakan untuk prapemprosesan data teks. Empat pengelas; NB, libSVM, KNN, dan pokok keputusan (J48) telah diuji dalam kajian ini. Ketepatan klasifikasi tertinggi sebanyak 94.5% dicapai pada masa jalan pengiraan 3.89 saat dengan teknik IG-CFS telah direkodkan di akhir kajian.

Ulumudin, et. al (2019) membuat penyelidikan bagi mengkaji cara membuat pengelasan data berbilang label iaitu topik ayat Al-Quran dengan pengelas KNN untuk memudahkan carian maklumat di dalam Al-Quran mengikut topik tertentu yang telah diklasifikasikan.

Perbandingan kejituan kajian dibandingkan dengan dua pengekstrakan ciri iaitu TF-IDF berwajaran dan TF-IDF. TF-IDF berwajaran digunakan dalam kajian ini untuk mengelakkan dominasi dokumen dengan sedikit jumlah istilah dalam query tetapi mempunyai frekuensi yang tinggi. Hasil daripada perbandingan itu, TF-IDF berwajaran mempunyai prestasi yang lebih baik berbanding TF-IDF biasa dengan keputusan skor-k paling optimum ialah $k=25$ dan purata skor kerugian hamming 0.134875.

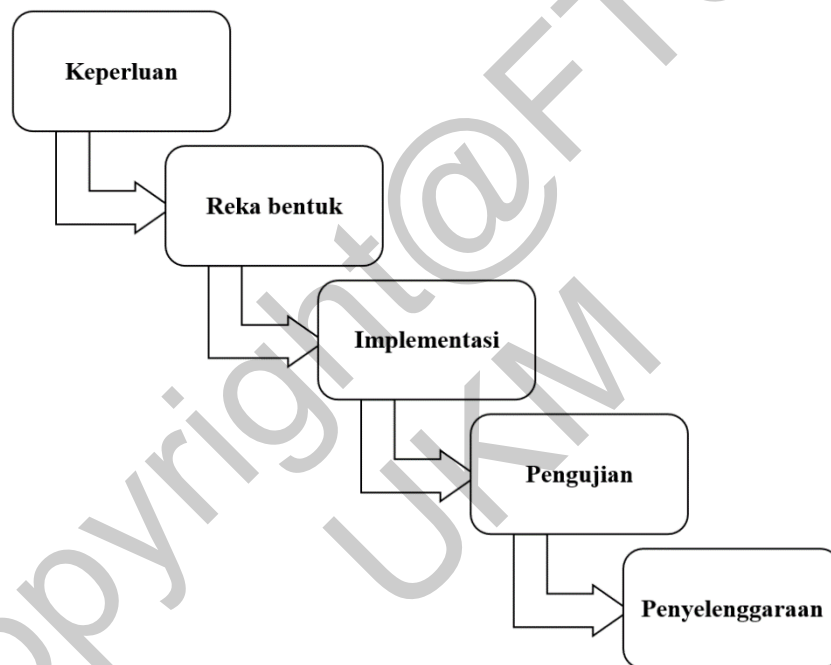
Hassan, et. al (2015) menjalankan kajian tentang Pengkategorian Tafsir Al-Quran menggunakan Pembelajaran Mesin KNN. Pengkategorian teks ialah proses pelabelan teks bahasa tabii kepada satu atau beberapa kategori daripada set yang telah ditetapkan. Objektif penyelidikan adalah, untuk merumuskan adakah algoritma KNN sesuai untuk pengelasan automatik dan sesuai bagi klasifikasi tafsir Al-Quran. Klasifikasi tersebut merangkumi pengenalpastian kategori tafsir Al-Quraan dalam bentuk kelas nombor dalam terjemahan Bahasa Melayu. Fasa pertama ialah prapemprosesan teks Arab dan kemudian menukar perkataan dalam Bahasa Arab kepada perkataan Melayu. Selepas itu, kategorikan kelas berdasarkan persamaan kosinus antara dokumen ujian dan dokumen latihan khusus. Hasilnya menunjukkan algoritma KNN adalah salah satu algoritma terbaik untuk pengkategorian tafsir Al-Quran berdasarkan prestasi kejituan yang tinggi dan kadar ralat yang rendah di akhir kajian.

Ahmed & Omar (2014) mengkaji hubungkait kaedah pemilihan ciri kepada peningkatan prestasi klasifikasi sentimen Melayu. Tiga pengelas pembelajaran mesin yang diselia dan tujuh kaedah pemilihan ciri digunakan untuk menentukan kaedah yang sesuai untuk klasifikasi ulasan tulisan Melayu dalam talian menggunakan analisis sentimen. Dapatan kajian menunjukkan bahawa klasifikasi sentimen Melayu bertambah baik menggunakan pendekatan pemilihan ciri. Ini menunjukkan bahawa semua kaedah pengurangan ciri secara amnya meningkatkan prestasi pengelas. Pendekatan SVM memberikan prestasi ketepatan tertinggi untuk pemilihan ciri bagi mengklasifikasikan sentimen Melayu berbanding dengan pendekatan

klasifikasi lain seperti PCA dan Chi-squared. SVM merekodkan 87% sebagai hasil ketepatan pemilihan ciri.

5) Metodologi Kajian

Bagi mencapai matlamat kajian, terdapat beberapa langkah utama yang perlu dilakukan secara berperingkat. Metodologi yang digunakan dalam kajian ini adalah berdasarkan model air terjun. Rajah 1 dibawah menunjukkan metodologi model air terjun yang akan dijalankan.



Rajah 1 Fasa-fasa dalam Metodologi Model Air Terjun

a) Keperluan

Dalam fasa ini melibatkan pengumpulan maklumat tentang fungsi yang diinginkan, jangkaan pengguna, kekangan dan sebarang spesifikasi lain yang berkaitan bagi menjalankan pengelasan, Hasil daripada fasa ini ialah dokumen keperluan komprehensif yang berfungsi sebagai panduan untuk proses pembangunan.

b) Reka bentuk

Setelah keperluan ditetapkan, reka bentuk seni bina peringkat tinggi pengelasan akan dibina untuk sistem. Ini termasuklah keseluruhan struktur, modul, antara muka dan aliran pengelasan.

c) Implementasi

Dalam fasa ini, pengekodan dan pengaturcaraan sebenar sistem berlaku. Pembangunan akan mengikut spesifikasi reka bentuk untuk menulis kod, mencipta modul dan menyepadukan model model pembelajaran mesin yang berbeza. Fasa implementasi memberi tumpuan kepada perubahan reka bentuk sistem yang dirancang sesuai dengan sistem yang ingin dibangunkan.

d) Pengujian

Setelah implementasi selesai, sistem akan memasuki fasa pengujian. Pengujian dilakukan untuk mengenalpasti kekurangan yang terdapat pada sistem. Pelbagai teknik ujian dijalankan untuk memastikan sistem pengelasan berfungsi dengan baik tanpa sebarang isu bagi penggunaan pengguna.

e) Penyelenggaraan

Semasa di fasa ini, sebarang isu yang timbul dalam sistem akan diselesaikan. Kemaskini juga juga boleh dibuat untuk menambahkan ciri baharu sesuai dengan keperluan pengguna yang berubah-ubah.

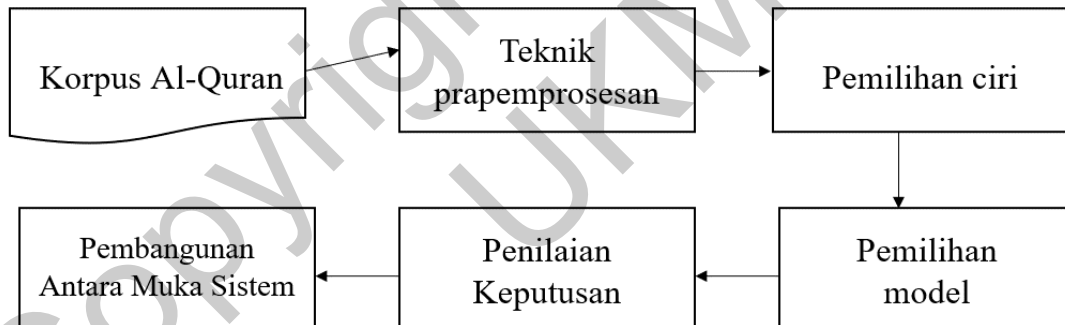
5) Organisasi Laporan

Laporan ini mengandungi lima bahagian di mana bahagian pertama iaitu bahagian pengenalan akan memberikan gambaran menyeluruh tentang kajian yang dijalankan. Seterusnya bahagian dua iaitu rangka kerja penyelidikan. Bahagian ini menerangkan proses pembangunan kajian ini berdasarkan rangka kerja yang dibangunkan. Rangka kerja kajian adalah penting kerana ia memberikan panduan dan penjelasan mengenai kajian yang telah dijalankan dan analisis data. Seterusnya bahagian tiga iaitu keputusan eksperimen dan perbincangan. Bahagian ini memaparkan hasil kajian dan maklumat yang diperolehi serta memberi makna dan kesimpulan

kepada kajian yang telah dijalankan. Seterusnya bahagian empat iaitu pembangunan antara muka sistem. Bahagian ini akan menerangkan pembangunan antara muka berdasarkan hasil kajian dengan teliti. Akhir sekali, bahagian lima iaitu kesimpulan. Bahagian ini merangkumi gambaran terhadap hasil kajian yang diperoleh. Bahagian ini menyimpulkan keseluruhan kajian dan memberi ringkasan tentang hasil kajian serta implikasinya.

Rangka Kerja Penyelidikan

Kajian ini dibangunkan berdasarkan rangka kerja penyelidikan yang mudah untuk difahami dan digunakan. Dengan menggunakan kaedah ini, setiap fasa harus dilengkapkan sebelum fasa seterusnya dimulakan. Di penghujung setiap fasa, penilaian dibuat bagi memastikan projek berjalan seperti yang dirancang. Rajah 2 menunjukkan aliran rangka kerja penyelidikan yang dijalankan.



Rajah 2 Rangka Kerja Penyelidikan

a) Korpus Al-Quran

Korpus yang digunakan dalam kajian ini telah mempunyai data yang dianotasi secara manual kepada dua kelas utama iaitu Al-Targhib (galakan) dan Al-Tarhib (amaran) berdasarkan kajian Hateem et. al (2018) iaitu pada teks Al-Quran dalam bahasa Arab. Korpus ini merangkumi 2000 data. Masing-masing terdiri daripada 1000 ayat galakan dan 1000 ayat amaran.

b) Teknik Prapemrosesan

Langkah prapemrosesan adalah penting dalam mencapai ketepatan yang tinggi dalam analisis dan ia bergantung pada teknik yang digunakan. Dalam kajian ini, data akan disediakan merangkumi proses memuatkan data dan melaksanakan prapemrosesan asas seperti penormalan, tokenisasi, serta lemmatisasi. Korpus kemudiaannya dibahagikan kepada set latihan dan set pengujian.

c) Pemilihan Ciri

Beberapa ciri akan digunakan bagi mendapatkan gabungan terbaik model dan ciri untuk menghasilkan pengelasan terbaik. Antara ciri yang digunakan ialah BOW, Word2Vec, TF-IDF dalam pemendaman kata.

d) Pemilihan Model

Pemilihan model akan melatih model pembelajaran mesin pada data berlabel. Dalam kajian ini Mesin Sokongan Vektor, *Random Forest*, *Naïve Bayes*, *Decision Tree*, *Logistic Regression*, *K-Nearest Neighbor* akan digunakan memandangkan model-model ini mampu mengendalikan set data yang besar dan kompleks.

e) Penilaian Keputusan

Penilaian kajian akan dilakukan berdasarkan penilaian ukuran setiap pengelasan binari menggunakan teknik yang berbeza. Pengelasan binari yang terbaik akan memberikan keputusan yang hampir sempurna. Metrik pengukuran capaian semula, kejituan dan pengukuran-F akan digunakan dalam menilai keputusan dan mendapatkan model dan ciri terbaik dalam pengelasan ayat Al-Quran kepada Al-Tarhib (galakan) dan Al-Tarhib (amaran).

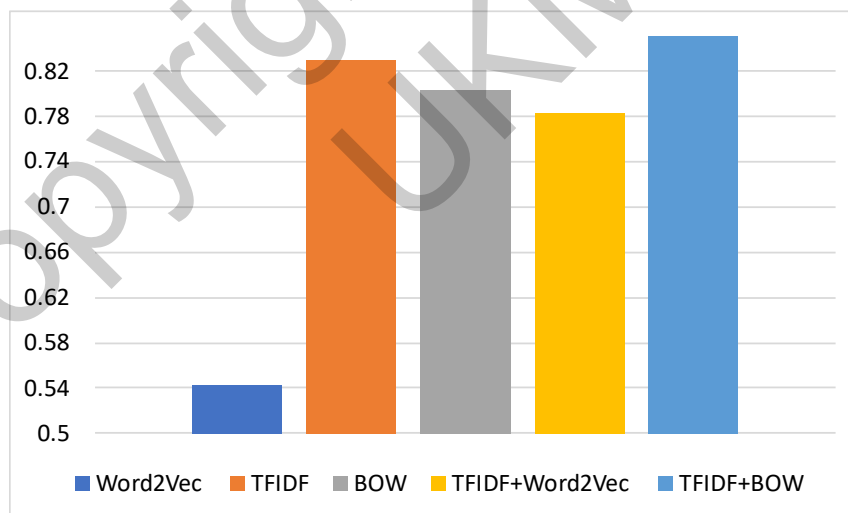
f) Pembanguna Antara Muka Sistem

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 nyata. Oleh itu, kajian ini telah menggunakan “Streamlit” untuk membangunkan aplikasi web. *Streamlit* adalah perisian *Python* atas talian yang boleh mencipta komponen UI yang boleh disesuaikan untuk model pembelajaran mesin serta sebarang API dan sebarang fungsi. Dengan adanya aplikasi web, objektif kajian ini dapat dicapai.

Keputusan Eksperimen dan Perbincangan

Dalam kajian ini, pelbagai model pembelajaran mesin dan ciri telah digunakan untuk pengelasan. Keputusan eksperimen yang diperoleh akan dianalisis seterusnya ciri dan model terbaik akan ditentukan berdasarkan keputusan tersebut.

1) Penentuan Ciri Terbaik



Rajah 3 Perbandingan Purata bagi Setiap Ciri

Rajah 3 menunjukkan bahawa ciri gabungan cenderung untuk mencapai ketepatan pengelasan yang lebih baik daripada menggunakan ciri tunggal. Berdasarkan keputusan eksperimen, ciri TF-IDF adalah ciri tunggal terbaik dan ciri gabungan TF-IDF + BOW

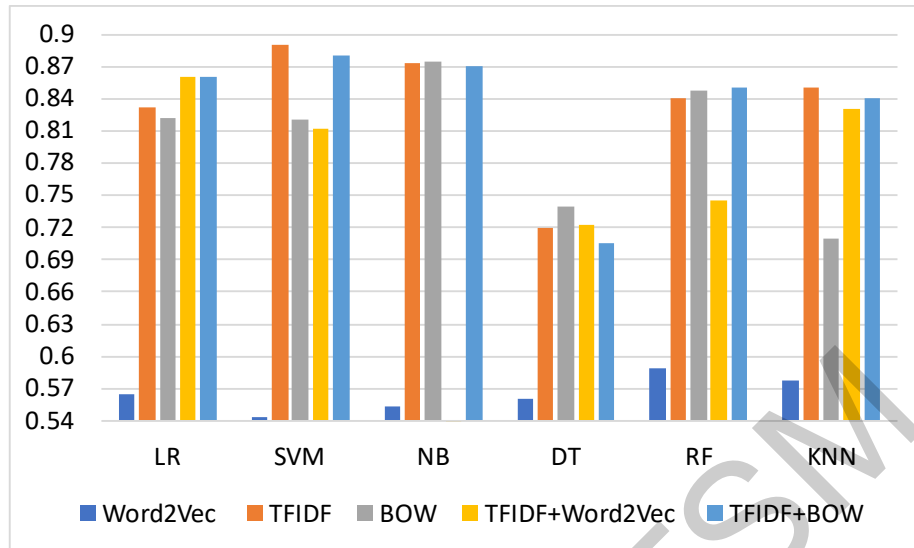
merupakan ciri gabungan terbaik. Secara keseluruhannya, TF-IDF + BOW merupakan ciri yang mendapatkan ketepatan pengelasan yang paling tinggi berbanding ciri-ciri lain yang digunakan dalam kajian ini dengan purata ketepatan **83.42%** dan nilai pengukuran-F1 sebanyak **83.41%**.

2) Penentuan Model Pengelasan Terbaik

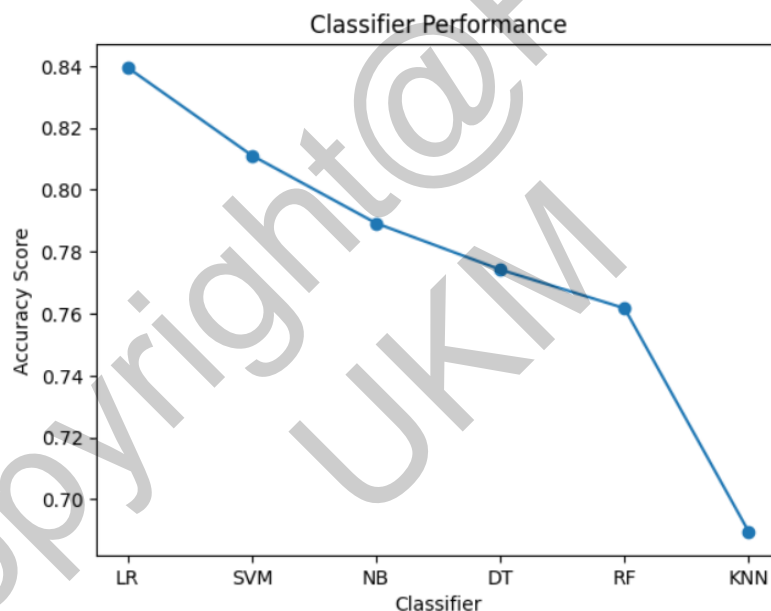
Model pengelas terbaik akan ditentukan berdasarkan ketepatan yang diperolehi oleh model tersebut berdasarkan ciri-ciri yang telah digunakan. Metrik penilaian lain seperti kejitian, skor capaian semula dan pengukuran-F1 turut dipertimbangkan untuk menentukan model yang terbaik. Perbandingan purata ketepatan model pengelas ditunjukkan dalam Jadual 2 di bawah.

Model	Ketepatan (Accuracy)	Kejitian (Precision)	Skor capaian semula (Recall)	Pengukuran-F1 (F1-score)
LR	83.95%	79.02%	78.94%	78.94%
SVM	81.11%	79.53%	78.95%	78.24%
NB	78.92%	79.44%	79.01%	78.22%
DT	77.42%	68.94%	69.61%	69.58%
RF	76.17%	76.37%	76.13%	76.11%
KNN	68.95%	75.60%	75.54%	75.52%

Jadual 2 Perbandingan Purata Ketepatan Model Pengelas



Rajah 4 Perbandingan Ketepatan Model Pengelas



Rajah 5 Perbandingan Purata Ketepatan Model Pengelas Menggunakan Graf Garisan

Berdasarkan Rajah 4 dan Rajah 5, jelas menunjukkan terdapat perbezaan ketepatan yang agak ketara bagi model pengelas yang digunakan. LR mencatatkan ketepatan pengelasan yang tertinggi secara purata iaitu sebanyak **83.95%** diikuti NB (**81.11%**), SVM (**78.92%**), RF (**77.42%**), KNN (**76.17%**) dan DT (**68.95%**). Manakala purata nilai pengukuran-F1 bagi setiap model masing-masing adalah LR (**78.94%**), NB (**78.22%**), SVM (**78.24%**), RF (**76.11%**), KNN (**75.52%**) dan DT (**69.58%**). LR memperoleh purata ketepatan pengelasan

dan purata nilai pengukuran-F1 yang lebih tinggi berbanding model pengelasan yang lain kerana model ini sesuai digunakan untuk pengelasan binari. LR berfungsi dengan lebih baik apabila apabila data dipisahkan secara linear dan kelas dipisahkan dengan baik sesuai dengan matlamat kajian ini iaitu membuat pengelasan Al-Quran kepada konsep Al-Tarhib (amaran) dan Al-Targhib (galakan). Prestasi setiap model berbeza-beza disebabkan oleh saiz set data, ketidakseimbangan kelas, hingar dan kehadiran *outlier*. Keseluruhannya, model pengelasan terbaik dalam kajian ini adalah LR.

3) Analisis Keputusan

Berdasarkan kajian yang dijalankan oleh Hateem et. al. (2018) iaitu mengenai pengelasan Al-Quran menggunakan analisis sentimen dalam Bahasa Arab klasik, kajian tersebut telah memperoleh keputusan ketetapan yang lebih baik berbanding kajian ini iaitu mencapai sehingga **93.95%** ketepatan dan nilai pengukuran-F1 sebanyak **93.65%** (rujuk Jadual 2). Antara faktor yang menyebabkan ketepatan yang rendah bagi kajian ini adalah disebabkan ayat-ayat Al-Quran telah diterjemahkan dalam Bahasa Inggeris yang berkemungkinan mengubah sedikit makna dan struktur ayat berbanding kajian yang dijalankan oleh Hateem yang menggunakan Bahasa Arab klasik. Hal ini kerana, terjemahan Al-Quran dalam Bahasa Inggeris adalah bergantung kepada penterjemahnya di mana setiap penterjemah mempunyai cara terjemahannya yang tersendiri. Bagi menangani isu ini, beberapa penterjemah diperlukan untuk melakukan terjemahan agar mendapat terjemahan ayat-ayat Al-Quran dengan tepat.

feature	# feature	Precision		Recall		F1	Accuracy
		encouragement	warning	encouragement	warning		
Lemma (baseline)	2826	76.94%	78.72%	79.40%	76.20%	78.15%	77.80%
N-Gram +POS	1000	97.33%	83.23%	80.30%	97.80%	88.00%	89.05%
N-Gram +POS	2000	97.81%	86.58%	84.80%	98.10%	90.84%	91.45%
N-Gram +POS	3000	98.56%	90.14%	89.20%	98.70%	93.65%	93.95%
POS	4000	97.58%	89.72%	88.80%	97.80%	92.98%	93.30%
N-Gram +POS	5000	97.58%	89.64%	88.70%	97.80%	92.93%	93.25%
N-Gram +POS	6000	97.57%	89.48%	88.50%	97.80%	92.81%	93.15%

Jadual 3 Keputusan Eksperimen Menggunakan N-Gram, Penandaan POS dan Pemilihan Ciri pada Pengelas NB berdasarkan Kajian Hateem et. al. (2018)

Berdasarkan Jadual 3, Hateem menjalankan eksperimen menggunakan Penandaan POS, model N-Gram dan pemilihan ciri berdasarkan korelasi. Ekperimen tersebut menunjukkan ketepatan **77.80%**, **89.05%**, **91.45%**, **93.95%**, **93.30%**, **93.25%** dan **93.15%** dicapai untuk pemilihan ciri dan ciri lemma dengan N-gram dan Penandaan POS (lemma 2826, 1000, 2000, 3000, 4000, 5000 dan 6000 masing-masing) untuk pengelas NB.

Di samping itu, dalam kajian ini menunjukkan gabungan ciri dan model terbaik iaitu ciri TF-IDF+BOW dan model LR memperoleh keputusan pengelasan tertinggi iaitu sebanyak **86%**. Keputusan yang diperoleh ditunjukkan dalam Jadual 4 di bawah.

Model	Ketepatan (Accuracy)	Kejituan (Precision)	Skor ingat semula (Recall)	Pengukuran-F1 (F1-score)
LR	86%	86%	86.04%	86%

Jadual 4 Keputusan (TF-IDF + BOW) + LR

Pembangunan Antara Muka Sistem

Antara muka bagi kajian ini merupakan sebuah interpretasi daripada keputusan analisis kajian. Setelah keputusan kajian yang diperoleh, ciri dan model yang terbaik berdasarkan ketepatan tertinggi dijadikan pengelas utama bagi menjayakan antara muka pengelasan ini.

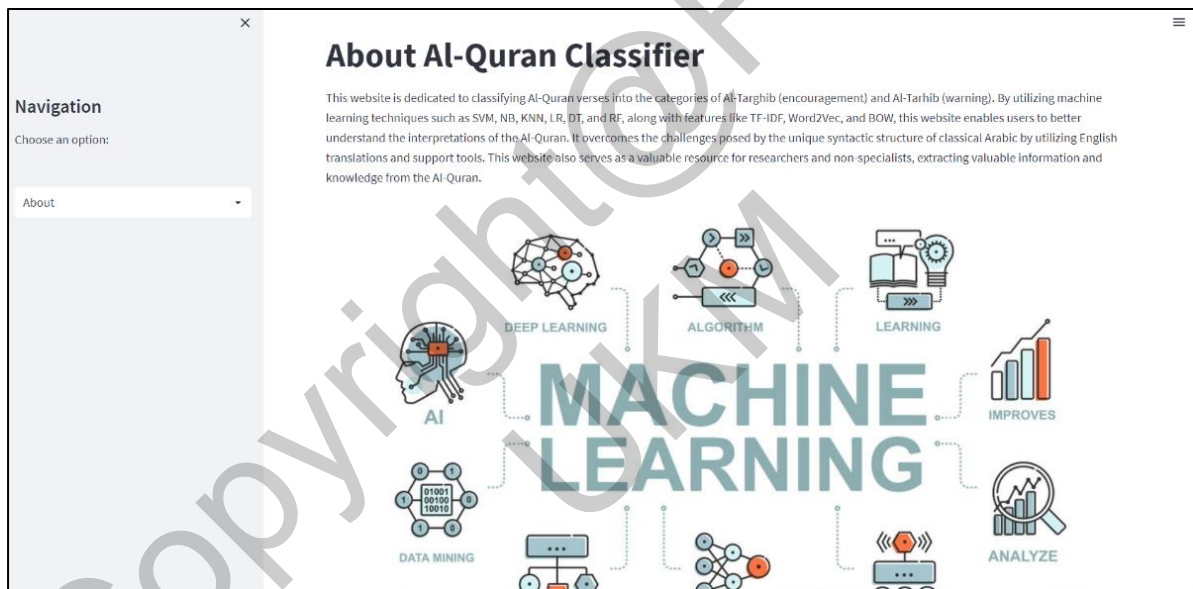
Bagi membina antara muka dalam kajian ini, bahagian belakang (*back end*) perlu disambungkan terlebih dahulu. Sambungan ini bertujuan untuk menyelaraskan analisis ciri dan model terbaik dengan antara muka pengelas yang dibangunkan. Pustaka “pickle” digunakan untuk mensiri struktur analisis kepada bentuk bait seterusnya menyimpannya dalam pangkalan data dalam dua fail iaitu fail “vectorizer.pkl” dan fail “model.pkl” (Rujuk Rajah 6).

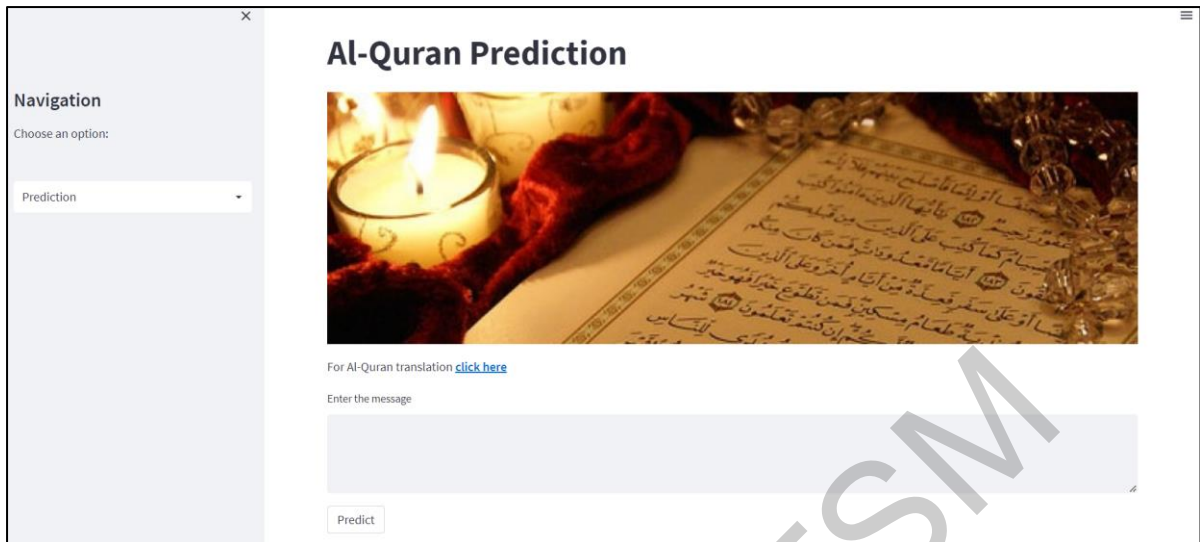
```
import pickle
pickle.dump(tfidf,open('vectorizer.pkl','wb'))
pickle.dump(mnb,open('model.pkl','wb'))
```

Rajah 6 Pustaka *pickle*

Pengaturcaraan bagi antara muka dimulakan dalam sebuah fail *Python* yang berasingan (*app.py*). Fail tersebut menyambungkan fail “vectorizer.pkl” dan “model.pkl” bersama-sama dengan perisian atas talian *Streamlit*. Setelah selesai membuat pengaturcaraan bagi antara muka tersebut, arahan “streamlit run app.py” di terminal akan melancarkan sistem antara muka yang telah dibangunkan.

Sistem ini mempunyai satu antara muka saja. Antara muka ini dibina menggunakan bahasa pengaturcaraan *Python*. Pengguna boleh menekan pilihan halaman di bahagian navigasi, “Navigation” di ruangan sebelah kiri halaman. Antara pilihan halaman yang disediakan adalah halaman “Home” iaitu halaman utama, halaman “About” yang menerangkan tentang matlamat kajian ini dan halaman “Prediction” iaitu halaman pengelasan. Reka bentuk halaman-halaman yang telah dinyatakan ditunjukkan dalam Rajah 7, Rajah 8 dan Rajah 9 di bawah.

Rajah 7 Antara Muka Halaman *Home*Rajah 8 Antara Muka Halaman *About*



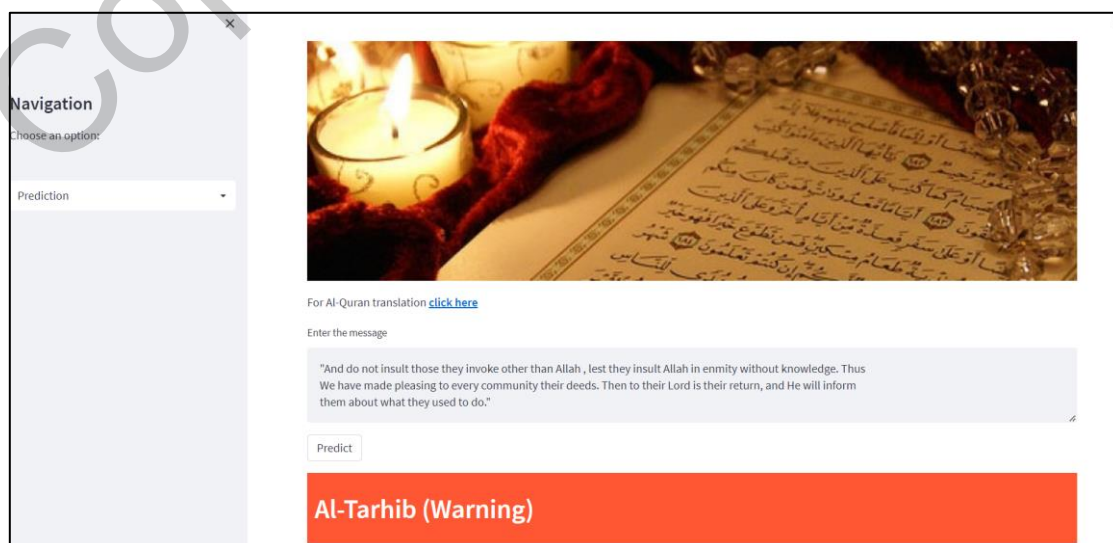
Rajah 9 Antara Muka Halaman *Prediction*

Pengguna boleh mula melakukan pengelasan dengan memasukkan potongan ayat Al-Quran dalam terjemahan Bahasa Inggeris di ruangan yang disediakan. Bagi mendapatkan ayat tersebut, pengguna boleh menekan URL yang disertakan. URL tersebut akan terus ke dokumen *Google Drive*, iaitu dokumen tersebut mempunyai terjemahan Al-Quran dalam Bahasa Inggeris (Rujuk Rajah 10). Seterusnya, pengguna boleh menyalin potongan ayat Al-Quran di laman web itu dan menampalnya di ruangan yang disediakan di antara muka pengelas Al-Quran.

Encouragement	Warning
<p>And give good tidings to those who believe and do righteous deeds that they will have gardens [in Paradise] beneath which rivers flow. Whenever they are provided with a provision of fruit therefrom, they will say, "This is what we were provided with before." And it is given to them in likeness. And they will have therein purified spouses, and they will abide therein eternally.</p>	<p>"And let not those who disbelieve ever think that [because] We extend their time [of enjoyment] it is better for them. We only extend it for them so that they may increase in sin, and for them is a humiliating punishment."</p>
<p>Indeed, as-Safa and al-Marwah are among the symbols of Allah. So whoever makes Hajj to the House or performs 'umrah - there is no blame upon him for walking between them. And whoever volunteers good - then indeed, Allah is appreciative and Knowing.</p>	<p>"But if you want to replace one wife with another and you have given one of them a great amount [in gifts], do not take [back] from it anything. Would you take it in injustice and manifest sin?"</p>
<p>They ask you, [O Muhammad], what they should spend. Say, "Whatever you spend of good is [to be] for parents and relatives and orphans and the needy and the traveler. And whatever you do of good - indeed, Allah is Knowing of it."</p>	<p>"And [also] those who spend of their wealth to be seen by the people and believe not in Allah nor in the Last Day. And he to whom Satan is a companion - then evil is he as a companion."</p>
<p>"That is from the news of the unseen which We reveal to you, [O Muhammad]. You knew it not, neither you nor your people, before this. So be patient; indeed, the [best] outcome is for the righteous."</p>	<p>"O you who were given the Scripture, believe in what We have sent down [to Muhammad], confirming that which is with you, before We obliterate faces and turn them toward their backs or curse them as We cursed the sabbath-breakers. And ever is the decree of Allah accomplished."</p>
<p>"And to Thamud [We sent] their brother Salih. He said, ""O my people, worship Allah; you have no deity other than Him. He has produced you from the earth and settled you in it, so ask forgiveness of Him and then repent to Him. Indeed, my Lord is near and responsive.""</p>	<p>"Wherever you may be, death will overtake you, even if you should be within towers of lofty construction. But if good comes to them, they say, ""This is from Allah""; and if evil befalls them, they say, ""This is from you."" Say, ""All [things] are from Allah. "" So what is [the matter] with those people that they can hardly understand any statement?"</p>
<p>So when Our command came, We saved Salih and those who believed with him, by mercy from Us, and [saved them] from the disgrace of that day. Indeed, it is your Lord who is the Powerful, the Exalted in Might.</p>	<p>"You will find others who wish to obtain security from you and [to] obtain security from their people. Every time they are returned to [the influence of] disbelief, they fall back into it. So if they do not withdraw from you or offer you peace or restrain their hands, then seize them and kill them wherever you overtake them. And those - We have made for you against them a clear authorization."</p>
<p>And each [story] We relate to you from the news of the messengers is that by which We make firm your heart. And there has come to you, in this, the truth and an instruction and a reminder for the believers.</p>	<p>"Paradise is not [obtained] by your wishful thinking nor by that of the People of the Scripture. Whoever does a wrong will be recompensed for it, and he will not find beside Allah a protector or a helper."</p>
<p>And thus will your Lord choose you and teach you the interpretation of narratives and complete His favor upon you and upon the family of Jacob, as He completed it upon your fathers before, Abraham and Isaac. Indeed, your Lord is Knowing and Wise.</p>	<p>"And from those who say, ""We are Christians"" We took their covenant, but they forgot a portion of that of which they were reminded. So We caused among them animosity and hatred until the Day of Resurrection. And Allah is going to inform them about what they used to do."</p>
<p>And he raised his parents upon the throne, and they bowed to him in prostration. And he said, "O my father, this is the explanation of my vision of before. My Lord has made it reality. And He was certainly good to me when He took me out of prison and brought you [here] from bedouin life after Satan had induced [estrangement] between me and my brothers. Indeed, my Lord is Subtle in what He wills. Indeed, it is He who is the Knowing, the Wise."</p>	<p>"And thus We have made for every prophet an enemy - devils from mankind and jinn, inspiring to one another decorative speech in delusion. But if your Lord had willed, they would not have done it, so leave them and that which they invent."</p>
<p>It is Allah who erected the heavens without pillars that you [can] see; then He established Himself above the Throne and made subject the sun and the moon, each running [its course] for a specified term. He arranges [each] matter; He details the signs that you may, of the meeting with your Lord, be certain.</p>	<p>"And do not insult those they invoke other than Allah; lest they insult Allah in enmity without knowledge. Thus We have made pleasing to every community their deeds. Then to their Lord is their return, and He will inform them about what they used to do."</p>

Rajah 10 Dokumen Terjemahan Al-Quran dalam Bahasa Inggeris

Setelah menampal potongan ayat tersebut, pengguna boleh menekan butang “Predict” untuk melakukan pengelasan. Sistem akan mengira ketepatan pengelasan berdasarkan potongan ayat tersebut seterusnya memaparkan keputusan pengelasan sama ada “Al-Tarhib (Encouragement)” atau “Al-Tarhib (Warning)” seperti yang ditunjukkan dalam Rajah 11 di bawah.



Rajah 11 Antara Muka Halaman Prediction Selepas Pengelasan

Rajah 11 menunjukkan antara muka sistem yang akan dibangunkan selepas membuat pengelasan. Teks yang dimasukkan dikategorikan sebagai konsep Al-Tarhib (Amaran) pada bahagian bawah teks. Melalui antara muka ini, pengguna dapat melihat paparan pengelasan konsep Al-Tarhib (galakan) dan Al-Tarhib (amaran) dengan lebih jelas. Pengguna juga boleh memasukkan potongan ayat baru di ruangan teks tersebut untuk membuat pengelasan teks baru.

Kesimpulan

Secara keseluruhannya, pembangunan web aplikasi terintegrasi model pembelajaran mesin telah disiapkan dalam tempoh masa yang telah ditetapkan. Pembangunan ini juga menepati objektif kajian yang telah disasarkan. Sistem ini dengan tujuan untuk memanfaatkan pengguna dapat mengklasifikasikan dan menganalisis kalamullah Al-Quran kepada konsep Al-Tarhib (galakan) dan Al-Tarhib (amaran). Keperluan sistem dan reka bentuk sistem telah dijelaskan dengan nyata.

1) Ringkasan Hasil Kajian

Bagi menjalankan kajian ini, korpus Al-Quran yang seimbang telah dianotasi secara manual daripada Al-Quran. Ia terdiri daripada 2000 ayat dimana 1000 daripadanya adalah ayat galakan dan 1000 yang lain adalah ayat amaran. Kajian ini menjadikan kajian daripada Hateem et. al. (2018) yang menjalankan pengelasan Al-Quran dalam Bahasa Arab dijadikan sebagai penanda aras untuk membandingkan beberapa eksperimen pada model pembelajaran mesin yang digunakan dalam kajian ini. Kajian ini telah mengimplementasi tiga ciri dan enam model pengelas untuk membuat pengelasan. Selain itu, metrik penilaian seperti ketepatan, kejituan, skor capaian semula dan pengukuran-F1 digunakan untuk menilai prestasi ciri dan model yang dipilih. Bagi menambahbaik ketepatan yang diperolehi, gabungan beberapa ciri seperti TF-IDF, Word2Vec dan BOW juga telah dikaji. Keputusan kajian menunjukkan bahawa ciri gabungan

TF-IDF+BOW dan model Logistic Regression mencapai ketepatan yang paling tinggi justeru menjadikan kedua-duanya sebagai ciri dan model yang terbaik dalam kajian ini.

2) **Impak dan Implikasi Kajian**

Algoritma mesin pembelajaran yang dapat memberikan ketepatan yang paling tinggi akan dipilih untuk pembangunan antara muka sistem aplikasi web. Pengekstrakan ciri seperti ciri gabungan juga telah menambahkan lagi prestasi pengklasifikasian sistem. Oleh itu, prestasi sistem yang dibangunkan dalam projek ini adalah amat baik. Selain itu, antara muka bagi kajian ini sangat mudah digunakan oleh semua pengguna kerana pengguna boleh melakukan pengelasan berulang kali tanpa had.

Kelebihan kajian ini adalah memberikan pengetahuan tentang pengelasan Al-Quran dalam terjemahan Bahasa Inggeris. Melalui kajian ini, dapatlah diketahui bahawa ciri dan teknik pengelasan sangat memainkan peranan dalam mendapatkan ketepatan yang baik dalam pengelasan. Dapat juga diteliti bahawa kalamullah Al-Quran memerlukan perhatian yang rapi untuk menganalisisnya agar setiap konteks dan makna tidak berubah semasa melakukan kajian ini. Di akhir kajian ini, model pengelas terbaik telah dikenalpasti dan sistem web pengelasan telah dibina. Sistem web tersebut membolehkan pengguna membuat pengelasan ayat-ayat Al-Quran dalam terjemahan Bahasa Inggeris.

3) **Kekurangan Kajian dan Cadangan Penambahbaikan Masa Depan**

Di samping itu, kajian ini juga mempunyai kekurangan. Kekurangan yang pertama adalah memperoleh ketepatan yang sedikit rendah berbanding kajian Hateem et. al. (2018) yang dijadikan penanda aras. Faktor-faktor yang membawa kepada kekurangan ini telah dibincangkan di Bahagian 6. Selain itu, sistem web yang dibangunkan adalah kurang menarik. Hal ini demikian kerana limitasi kepada perisian yang digunakan akibat kekurangan dana.

Sistem web yang lebih mesra pengguna dan menarik dapat dibangunkan sekiranya masalah dana ini dapat diatasi.

Pada kajian masa depan, pendekatan pembelajaran mendalam (*deep learning*) akan cuba diimplementasikan untuk memperoleh ketepatan yang lebih baik. Selain itu, kaedah pemendaman kata seperti “BERT” akan digunakan kerana “BERT” berupaya untuk memahami konteks ayat dan mengendalikan pengelasan teks dengan sangat baik berbanding kebanyakan kaedah pemendaman kata yang lain. BERT juga menyokong data yang berbilang bahasa justeru dapat memanfaatkan pengetahuan tersebut untuk melakukan klasifikasi dengan lebih baik. Akhir sekali, diharapkan kajian ini boleh diperluaskan lagi dalam bidang Hadis dan Sunnah supaya lebih ramai lagi akan memperoleh manfaat daripadanya.

Penghargaan

Terlebih dahulu, saya ingin mengambil peluang ini untuk mengucapkan jutaan terima kasih kepada penyelia saya, Dr. Saidah binti Saad yang telah sabar membimbing dan memberi tunjuk ajar serta ilmu yang berkenaan kepada saya sepanjang kajian ilmiah dijalankan. Jutaan terima kasih juga saya hendak ucapkan kepada para pensyarah dari FTSM yang telah menaburkan ilmu pengetahuan kepada saya sepanjang pengajian saya di Universiti Kebangsaan Malaysia (UKM).

Selain itu, saya juga ingin mengucapkan penghargaan terima kasih yang tidak terhingga kepada ibu bapa saya Encik Azman Bin Sulong dan Puan Jamilah Binti Kasran yang selalu menjadi sumber inspirasi, memberi semangat dan sentiasa berdoa terhadap kejayaan saya dalam menyiapkan kajian ini. Tidak lupa juga kepada sahabat seperjuangan Abdul Azim, Nur Farazila, Mimy Sarrah, Arvind Naido dan Adam Haikal yang banyak membantu saya dalam memberikan pandangan yang bernas serta sokongan kepada saya semasa diperlukan. Hanya Allah SWT yang mampu membalas jasa kalian.

Rujukan

- Adeleke, A., Samsudin, N., Mustapha, A., & Khalid, S. A. (2018). Automating quranic verses labeling using machine learning approach. *Indonesian Journal of Electrical Engineering and Computer Science*, 10(1), 925-931.
- Adeleke, A., & Samsudin, N. (2018). A Hybrid Feature Selection Technique for Classification of Group-based Holy Quran Verses. *International J of Engineering & Technology*, 7, 228-233.
- Aghahadi, Z., 2018. Word Embedding In Small Corpora: A Case Study in Quran. In:2018 8th International Conference on Computer and Knowledge Engineering (ICCCKE), (Iccke).pp. 303–307.
- AlHasani, H., Saad, S., & Kassim, J. (2018). Classification of Encouragement (Targhib) And Warning (Tarhib) Using Sentiment Analysis on Classical Arabic. *International Journal on Advanced Science, Engineering and Information Technology*, 8(4-2), 8-4.
- Alsaffar, A., & Omar, N. (2014, November). Study on feature selection and machine learning algorithms for Malay sentiment classification. In *Proceedings of the 6th International Conference on Information Technology and Multimedia* (pp. 270-275). IEEE.
- Hassan, G. S., Mohammad, S. K., & Alwan, F. M. (2015). Categorization of ‘Holy Quran-Tafseer’ using K-nearest neighbor algorithm. *Int. J. Comput. Appl*, 129(12), 1-6.
- Hira, Z. M. & Gillies, D. F. 2015. A review of feature selection and feature extraction methods applied on microarray data. *Advances in bioinformatics* 2015.
- Haleem, M. A. A. 1999. *Understanding the Qur’an: Themes and Styles*. London Qur’an studies. I. B. Tauris. Retrieved from https://books.google.com.my/books?id=KoCBBX_csvUC
- Kübler, S., Liu, C. & Sayyed, Z. A. 2017. To use or not to use: Feature selection for sentiment analysis of highly imbalanced data. *Natural Language Engineering* 1–35.

- Mourad, A. & Darwish, K. 2013. Subjectivity and Sentiment Analysis of Modern Standard Arabic and Arabic Microblogs
- Paltoglou, G. & Thelwall, M. 2010. A study of information retrieval weighting schemes for sentiment analysis. Proceedings of the 48th annual meeting of the association for computational linguistics, hlm. 1386–1395. Association for Computational Linguistics.
- Sallam, R. M., Mousa, H. M. & Hussein, M. 2016. Improving Arabic Text Categorization using Normalization and Stemming Techniques.
- Ulumudin, G. I., Adiwijaya, A., & Mubarak, M. S. (2019, March). A multilabel classification on topics of qur'anic verses in English translation using K-Nearest Neighbor method with Weighted TF-IDF. In Journal of Physics: Conference Series (Vol. 1192, No. 1, p. 012026). IOP Publishing.
- Yadollahi, A., Shahraki, A. G. & Zaiane, O. R. 2017. Current State of Text Sentiment
- Zeroual, I. & Lakhouaja, A. 2017. Arabic information retrieval: Stemming or lemmatization? 2017 Intelligent Systems and Computer Vision (ISCV), hlm. 1–6. IEEE.

Siti Nurizzati Binti Azman (A180777)
Dr. Saidah Binti Saad
Fakulti Teknologi & Sains Maklumat,
Universiti Kebangsaan Malaysia