

PROPOSING NLP ALGORITHMS FOR SHARIAH
SCREENING PLATFORM (SHARSP)

BY

MARHANUM CHE MOHD SALLEH

A dissertation submitted in fulfillment of the requirement for
the degree of Master of Computing

Kulliyyah of Information and Communication Technology
International Islamic University Malaysia

NOVEMBER 2023

ABSTRACT

This research proposes an algorithm for Shariah Screening Platform to assist Shariah decision making process in the Islamic Financial Industry. It is designed based on mixed methodology approach which involves both quantitative as well as qualitative method for data collection and analysis. Qualitative approach is first done by collecting sample of bank product proposal and guidelines from the Central Bank of Malaysia for similarity algorithm development. The first test of similarity is done using Levenshtein Distance. Given the algorithm written in Python, this research tests two different sets of text data where the first one is sample of Tawarruq product proposal (query) with Tawarruq policy document (corpus) and the second set is the same Tawarruq proposal (query) with Shariah Governance Policy Document as corpus. A suitable algorithm is built for document similarity analysis using Python programming language. Two measures of similarity is adopted which are Levenshtein Distance and FuzzyWuzzy Package to compare two strings document. These algorithms then are used to develop a Shariah Screening Platform named as SHARSP. The second stage is quantitative approach where data is collected through online structured survey distributed among industry practitioners to observe their view on the proposed Shariah Screening Platform. Based on Technology Acceptance Model (TAM) and Technology Readiness Index (TRI) as a measurement tool and multiple regression analysis, majority of the respondents were in view that the proposed SHARSP is useful, and they were prone towards the industry innovation. Overall, outputs of this research are significant to Islamic financial industry in proposing new technology for document screening. It would also add to the existing literatures on Islamic finance especially in the adoption of NLP technology.

Keywords: Shariah Screening Platform, Islamic Financial Industry, TAM, TRI, NLP

خلاصة

يقترح هذا البحث خوارزمية لمنصة فحص الشريعة للمساعدة في عملية صنع القرار الموافق للشريعة في الصناعة المالية الإسلامية. تم تصميم الخوارزمية باستخدام المنهجية المختلطة التي تتضمن طريقة كمية ونوعية لجميع البيانات وتحليلها. تم أولاً استخدام النهج النوعي من خلال جمع عينه من اقتراح المنتج المصرفي والإرشادات من البنك المركزي في ماليزيا لتطوير خوارزمية التشابه. حيث يتم إجراء الاختبار الأول للتشابه باستخدام مسافة ليفنشتاين Distance Levenshtein. بالنظر إلى الخوارزمية المكتوبة في Python، فإن هذا البحث يختبر مجموعتي مختلفتي من بيانات النص حيث تكون الأولى عينة من مقترح منتج التوارق (الاستعلام) مع وثيقة سياسة التوارق (الموسوعة)، والمجموعة الثانية هي نفس مقترح التوارق (الاستعلام) مع وثيقة سياسة حوكمة الشريعة كمصطلح. تام إنشاء خوارزمية مناسبة لتحليل تشابه الوثائق باستخدام لغة البرمجة Python. حيث تام اعتماد قياسين المتشابه هما مسافة ليفنشتاين وحزمة FuzzyWuzzy لمقارنة وثيقتي نصيتي. واستخدمت هذه الخوارزميات بعد ذلك لتطوير منصة فحص الشريعة بإسم (SHARSP) أما المرحلة الثانية فهي استخدام النهج الكمي، حيث تام جمع البيانات من خلال استبيان منظم عرب الإنترنت ثم توزيعه بين ممارسي الصناعة لمراقبة آرائهم حول منصة فحص الشريعة المقترحة. وبناء على نموذج قبول التكنولوجيا (TAM) ومؤشر استعداد التكنولوجيا (TRI) كأداة قياس وتحليل الانحدار المتعدد، كانت آراء غالبية المستجيبين أن منصة SHARSP المقترحة مفيدة، وكانوا يميلون للابتكار في الصناعة. و بشكل عام، تعتبر

هذا البحث مهمة لصناعة التمويل الإسلامي في اقتراح تكنولوجيا جديدة لفحص الوثائق. كما أنها ستسهم أيضا في الأدبيات الحالية حول التمويل الإسلامي خاصة في اعتماد تكنولوجيا معالجة اللغة الطبيعية (NLP)

الكلمات المفتاحية: منصة الفحص الشرعي، الصناعة المالية الإسلامية، TAM، TRI، NLP



APPROVAL PAGE

I certify that I have supervised and read this study and that in my opinion, it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Master of Computing.

.....
Dr. Rizal Mohd Nor
Supervisor

.....
Assoc. Prof. Dr. Adamu Abubakar Ibrahim
Co-Supervisor

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Master of Computing.

.....
Dr. Amir 'Aatieff Bin Amir Hussin
Internal Examiner

.....
Prof. Ts. Dr. Burhanuddin Bin Mohd Aboobaidar
External Examiner

This dissertation was submitted to the Department of Computer Science and is accepted as a fulfilment of the requirement for the degree of Master of Computing.

.....
Dr. Amir 'Aatieff Bin Amir Hussin
Head, Department of Computer
Science


This dissertation was submitted to the Kulliyah of Information & Communication Technology and is accepted as a fulfilment of the requirement for the degree of Master of Computing.

.....
Prof. Dr. Murni Bt. Mahmud
Dean, Kulliyah of Information &
Communication Technology

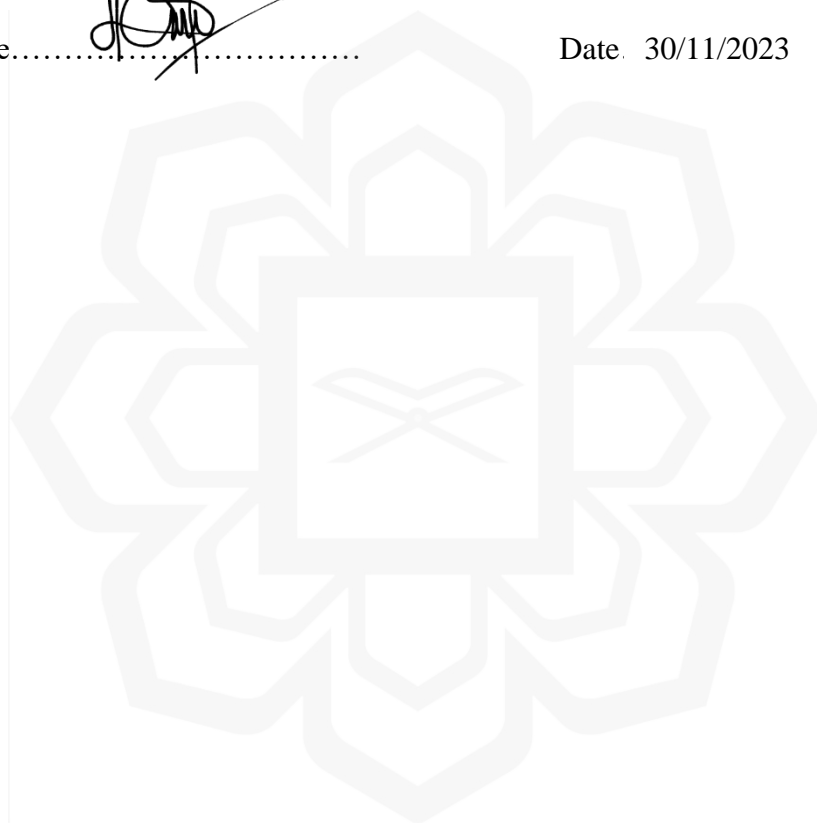
DECLARATION

I hereby declare that this dissertation is the result of my own investigations, except where otherwise stated. I also declare that it has not been previously or concurrently submitted as a whole for any other degrees at IIUM or other institutions.

Marhanum Che Mohd Salleh

Signature.....

Date 30/11/2023



INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

**DECLARATION OF COPYRIGHT AND AFFIRMATION OF
FAIR USE OF UNPUBLISHED RESEARCH**

TITLE OF THE THESIS/DISSERTATION

I declare that the copyright holder of this thesis/dissertation are jointly owned by the student and IIUM.

Copyright © 2014 Student Name and International Islamic University Malaysia. All rights reserved.

No part of this unpublished research may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without prior written permission of the copyright holder except as provided below

1. Any material contained in or derived from this unpublished research may only be used by others in their writing with due acknowledgement.
2. IIUM or its library will have the right to make and transmit copies (print or electronic) for institutional and academic purpose.
3. The IIUM library will have the right to make, store in a retrieval system and supply copies of this unpublished research if requested by other universities and research libraries.

By signing this form, I acknowledged that I have read and understand the IIUM Intellectual Property Right and Commercialization policy.

Affirmed by Marhanum Che Mohd Salleh



.....

Signature

30/11/2023

.....

Date

INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

**DECLARATION OF COPYRIGHT AND AFFIRMATION OF
FAIR USE OF UNPUBLISHED RESEARCH**

TITLE OF THE THESIS/DISSERTATION

I declare that the copyright holder of this thesis/dissertation is International Islamic University Malaysia.

Copyright © 2014 International Islamic University Malaysia. All rights reserved.

No part of this unpublished research may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without prior written permission of the copyright holder except as provided below

1. Any material contained in or derived from this unpublished research may only be used by others in their writing with due acknowledgement.
2. IIUM or its library will have the right to make and transmit copies (print or electronic) for institutional and academic purpose.
3. The IIUM library will have the right to make, store in a retrieval system and supply copies of this unpublished research if requested by other universities and research libraries.

By signing this form, I acknowledged that I have read and understand the IIUM Intellectual Property Right and Commercialization policy.

Affirmed by Marhanum Che Mohd Salleh



.....
Signature

30/11/2023
.....

Date

INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

**DECLARATION OF COPYRIGHT AND AFFIRMATION OF
FAIR USE OF UNPUBLISHED RESEARCH**

TITLE OF THE THESIS/DISSERTATION

I declare that the copyright holder of this thesis/dissertation is Name of the Student.

Copyright © 2014 Student Name. All rights reserved.

No part of this unpublished research may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without prior written permission of the copyright holder except as provided below

1. Any material contained in or derived from this unpublished research may only be used by others in their writing with due acknowledgement.
2. IIUM or its library will have the right to make and transmit copies (print or electronic) for institutional and academic purpose.
3. The IIUM library will have the right to make, store in a retrieval system and supply copies of this unpublished research if requested by other universities and research libraries.

By signing this form, I acknowledged that I have read and understand the IIUM Intellectual Property Right and Commercialization policy.

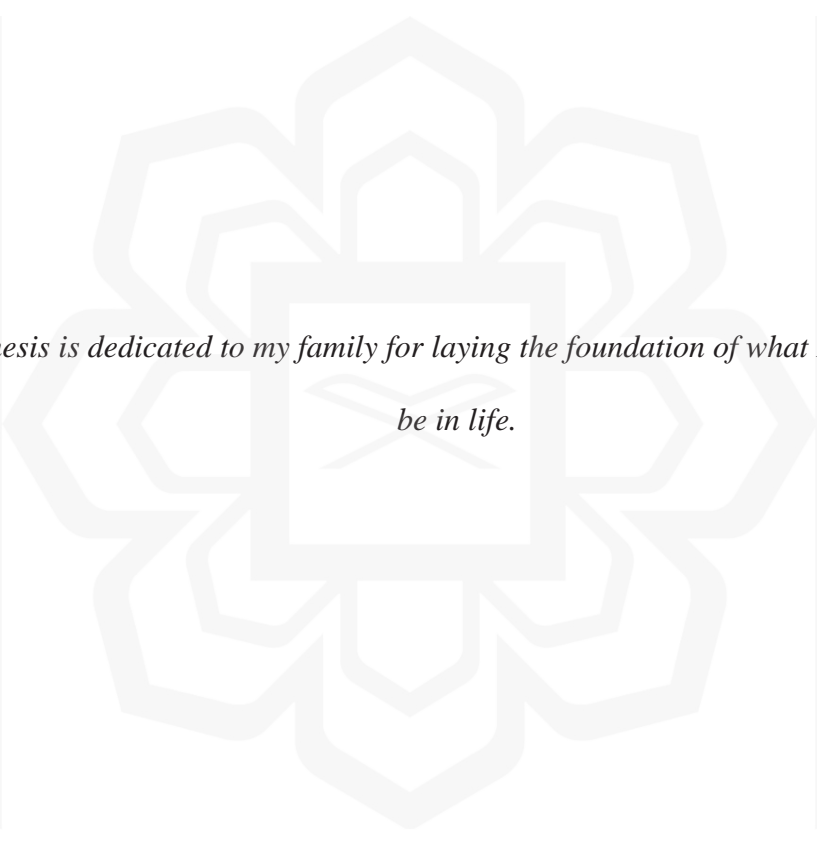
Affirmed by Marhanum Che Mohd Salleh



.....
Signature

30/11/2023

.....
Date



*This thesis is dedicated to my family for laying the foundation of what I turned out to
be in life.*

ACKNOWLEDGEMENTS

In the name of Allah, the Most Gracious and the Most Merciful.

Alhamdulillah, all praises to Allah for the strengths and His blessing in completing this thesis.

Special appreciation goes to my supervisor, Dr. Rizal Mohd Nor, for his supervision and constant support. His invaluable help of constructive comments and suggestions throughout the experimental and thesis works have contributed to the success of this research. Not forgotten, my appreciation to my co-supervisor, Assoc. Prof. Dr. Adamu Abubakar Ibrahim.

I would like to express my appreciation to the Dean and former Dean, Kulliyah of Information and Communication Technology, Dean Kulliyah of Economics and Management Sciences, and Head Department of Finance for their support and help towards my postgraduate affairs. My acknowledgement also goes to all colleagues, Department of Finance for their invaluable support.

Finally, my deepest gratitude goes to my beloved parents; my late father Che Mohd Salleh, my mother HjH Maimunah, my mother inlaw HjH Wan Noraini, and also to my siblings for their endless love, prayers and encouragement. Also not forgetting my special thanks to my beloved husband Riwan Ramli and lovely daughter Safiya, for their endless love and support. To those who indirectly contributed to this research, your kindness means a lot to me.

Thank you very much.

Wassalam.

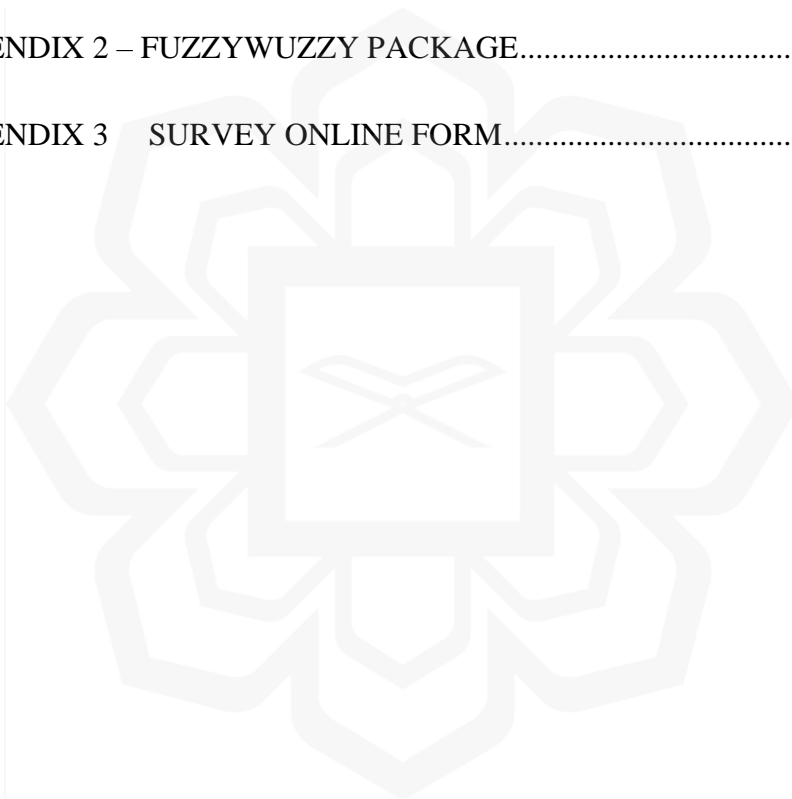
TABLE OF CONTENTS

ABSTRACT.....	ii
APPROVAL	ii
DECLARATION	vii
ACKNOWLEDGEMENTS	xii
TABLE OF CONTENTS.....	xiii
LIST OF TABLES	xvii
LIST OF FIGURES	xviii
CHAPTER ONE	1
INTRODUCTION	1
1.0 Research Background	1
1.1 Problem Statement	2
1.2 Research Questions	4
1.3 Research Objectives	5
1.4 Significance of Research.....	5
1.5 Chapter Summary	6
CHAPTER TWO	7
LITERATURE REVIEW	7
2.0 Introduction.....	7
2.1 Islamic Financial Industry (IFI)	7
2.1.1 Financial Technology (Fintech).....	8
2.1.2 Fintech and Banking Business	10
2.1.3 Shariah Decision Making Process in IFI	11

2.2	Information Retrieval	13
2.3	IR Retrieval Process.....	15
2.3.1	Dimensions of Similarity	16
2.3.1.1	Lexical Similarity.....	16
2.3.1.2	Structural Similarity.....	16
2.3.1.3	Semantic Similarity.....	17
2.4	IR Model	17
2.4.1	Exact Match Model.....	17
2.4.2	Vector Space Model (VSM)	18
2.4.3	Probabilistic models.....	21
2.5	Literatures on Natural Language Processing (NLP).....	22
2.5.1	Existing Works on NLP	24
2.6	Chapter Summary	29
CHAPTER THREE		30
RESEARCH METHODOLOGY.....		30
3.0	Introduction.....	30
3.1	Research Design.....	30
3.2	Proposed Algorithm	31
3.3	Data Preprocessing.....	32
3.4	Measure of Similarity	33
3.4.1	Levenshtein Distance	33

3.4.2 FuzzyWuzzy Package	35
3.5 Survey among Industry Practitioners.....	37
3.5.1 Technology Acceptance Model (TAM) and Technology Readiness Index (TRI)	38
3.5.2 Survey instrument	38
3.5.2 Sample Selection.....	40
3.5.3 Sample size	41
3.5.4 Survey Analysis	42
3.6 Chapter Summary	42
CHAPTER FOUR	43
RESULTS AND DISCUSSION	43
4.0 Introduction.....	43
4.1 Document Pre-Processing.....	43
4.2 Similarity Analyses.....	44
4.2.1 Levenshtein Distance	44
4.2.2 FuzzyWuzzy Package	47
4.4 Results of Survey	49
4.4.1 Descriptive Analysis	49
4.4.2 Multiple Regression Analysis	50
CHAPTER FIVE	54
CONCLUSION.....	54
5.0 Introduction.....	54

5.1 Discussion	54
5.2 Limitation.....	56
5.3 Suggestion for Future Research	57
5.4 Chapter Summary	57
REFERENCES	58
APPENDIX 1 - LEVENSHTEIN DISTANCE	65
APPENDIX 2 – FUZZYWUZZY PACKAGE.....	67
APPENDIX 3 SURVEY ONLINE FORM.....	68



LIST OF TABLES

Table 2. 1: Product Approval Process	12
Table 4.1: FuzzyWuzzy Ratios Value	47
Table 4.2: Gender	49
Table 4.3: Job Position	49
Table 4.4: Company	50
Table 4.5: Work Experience	50
Table 4.6: Model Summary	51
Table 4.7: ANOVA	51
Table 4.8: Results of Coefficients	52

LIST OF FIGURES

Figure 2.1: IR Process	15
Figure 2.2: Sample of a query	19
Figure 2.3: Number of literatures on NLP by year	25
Figure 2.4: Total number of NLP literatures by author	25
Figure 2.5: Countries that have contributed to NLP literatures	26
Figure 2.6: Literatures on NLP by subject area	26
Figure 2.7: NLP literatures on social sciences	27
Figure 2.8: NLP literatures from business, economics, and finance area	28
Figure 2.9: Main contributors on NLP literatures	28
Figure 3.1: Research Design	31
Figure 3.2: Proposed algorithm for document screening	32

CHAPTER ONE

INTRODUCTION

1.0 Research Background

Technology invention have started since few decades ago and it was the work of groups of scientists, programmers, and engineers. Among the pioneers of the technology are Paul Otlet and Vannevar Bush who have figured out online storage systems since 1930s and Nikola Tesla which have come out with the idea of wireless system in the early 1900s. Meanwhile, the first Internet prototype have been funded by U.S. Department of Defence in 1960 via the ARPANET (Advanced Research Projects Agency Network) project. It then became a modern internet when a computer scientist Tim Berners-Lee introduces World Wide Web (WWW) in year 1990 as the main objective is to access data online (www.history.com).

As huge of data was gathered online nowadays, difficulty of selecting a correct and reliable data have become a challenge. Furthermore, tendency to manipulate data have also increased among those who keen to take advantage over the technology advancement. Its now depends on human intention whether to use the data for the benefits of future or to bring into distraction of the world environment. If the data is used to give benefits to the planet, people would need to select which data is to be used and how it to be gathered. Computer technologies have proven that online information retrieval have become more efficient and able to reach users with more in-depth information than the traditional approach (full texts and hardcopy documents).

Hence, in the current information-age society, advanced retrieval systems and the automatic abstraction of information from streams of text, coupled with automated indexing, summarization and abstracting become the societal needs which is offered by Intelligent Information Retrieval (IIR). The marriage between IR and Artificial Intelligent (AI) systems are important in deriving huge quantities of raw text in an intelligent manner (Intelligent Information Retrieval). In addition, because most of the current databases are non-bibliographic databases and given that free style queries requested directly from innocent online searchers, the intelligent information retrieval (IIR), which is controlled by automatic, machine-performed procedures, is needed.

This research therefore in general is conducted to develop an algorithm for automatic assessment of Shariah document using Natural Language Processing (NLP) technology for document matching to assist Shariah decision making process for Islamic Financial Industry (IFI). It is believed that with the adoption and implementation of NLP in the IFI, it will give an efficient platform for the Shariah scholars as well as industry players in Shariah decision making process as well as to solve any Shariah issues.

1.1 Problem Statement

Digitalization is transforming how people interact and do daily business. The advancements in banking technology have influenced the future of financial services around the world. The establishment of digital banking has now in its growing stage where the way the financial services is offered to all is positively accepted especially by our new generation of digitally savvy millennials and Gen Zers. Through technology, the business has now is more towards business-to-customer (BTC) where the financial services are directly reached to the customer in a fingertip and within seconds. The usage of artificial intelligent (AI) has complimented the

roles of human as financial advisor and through blockchain technology, the whole financial system becomes efficient to tie all segments of banking businesses.

In spite of all the advantages that the technology has brought to the financial market, there are few segments that still need attention to be digitalised especially the Islamic banking business that require attention on Shariah compliancy in all aspects. It is observed that there is limitation of technology usage in the area of product structuring mainly on documentation where it is done manually by bank officers. Compared to the conventional products, the Islamic banking products require more documents based on the nature of contracts adopted. It is the duty of Shariah officers to ensure that all documents are comply with the Shariah and the central bank policy in the process of structuring new products. This duty would require a lot of effort, time, and energy to verify that the product will not having any element of Shariah non-compliant. Hence, the advancement of Natural Language Processing (NLP) technology would be necessary and significant to assist this process and reduce the burden of the Shariah officers.

It is also observed that there is limited number of Robo Advisor exist in the financial market and it only focused on digital investment management services. Based on observation, there are five Robo Advisors currently in Malaysian financial market which are MyTheo, StashAway, BEST, Wahed Invest, and Raiz. Unfortunately, none of these Robo Advisors have aided the Islamic banks with regards to the Shariah compliant status of the products during the structuring process. Based on suggestion by Sa'ad et al. (2020), the involvement of Robo advisory will ease the product structuring process by evaluating the sources of Shariah to provide necessary information for the physical Shariah advisor to accomplish the ruling accordingly. This effort may be started with a development of NLP algorithm for document matching or similarity screening.

Moreover, there are quite a number of Shariah non-compliant cases among Islamic banks in Malaysia where the root cause is on the documentation issues (Ali & Hassan, 2020; Mohamad & Ab Rahman, 2014; Radzi, 2012). As reported, most of the cases are related to the contract documentation, calculation of Ta'widh/Ibra', and term of conditions which contradict with the nature of contract (sample of cases: *CIMB Islamic Bank Bhd v LCL Corp Bhd & Anor [2012] 3 MLJ 869* and one Court of Appeal case of *Pan Northern Air Services Sdn Bhd v Maybank Islamic Bhd and another appeal [2021] 3 MLJ 408*). The existence of an efficient platform (mechanism) to retrieve relevant document efficiently would assist Shariah officer as well as the Shariah Committee to conduct earlier screening of Shariah status of the documents/products proposed.

Finally, it is a standard enforced by the Central Bank of Malaysia through Shariah Governance Policy Document which was issued on September 2019 for the Shariah committee to “establish a robust methodology to guide its decision-making process” (Standard 10.4) and they “shall be accountable for the quality, accuracy and soundness of its own decision or advice” (Standard 10.3). Therefore, it is believed that the proposed Shariah Screening Platform (SHARSP) using NLP technology would contribute significantly to the Islamic financial industry in introducing a digital platform for document screening. The research questions are the following.

1.2 Research Questions

1. What are the suitable algorithms for a digital assessment platform for Shariah document screening using Natural Language Processing (NLP)
2. What are the views of industry practitioners on the proposed Shariah Screening Platform named as SHARSP?

1.3 Research Objectives

1. To propose the suitable algorithms for a digital assessment of Shariah document using Natural Language Processing (NLP)
2. To investigate views of industry practitioners on the proposed Shariah Screening Platform named as SHARSP

1.4 Significance of Research

There are two important outputs of this research which first, algorithms for document screening based on Leveshtein Distance and FuqzzyWuzzy Package which is used to build a Shariah Screening Platform (SHARSP). This platform is currently in its infancy level and will be enhanced in future. It will become the first Shariah screening platform meant for internal use by Shariah officers and Shariah committee to check the document of product proposals whether it meet Bank Negara Policy or other standards. It would also reduce the level of stress and time consuming among practitioners in the process of product development and at the same time strengthen the compliancy of the industry towards the Shariah.

Second, this research provides views of the industry practitioners with a sample of 50 respondents in the survey conducted recently. Results from the survey is significant to show that the Shariah screening platform would be useful for the industry, and it will influence their intention to adopt this platform. Finally, this research is prominent in adding the existing literatures on financial technology (Fintech) as well as Islamic financial literatures. Findings of this research would serve as an initial effort to develop a Shariah screening platform based on NLP and later the usage of AI to enhance the practice of Islamic financial industry.

1.5 Chapter Summary

This chapter discusses the background of this research which includes problem statement, research questions, objectives, and significance of research. It is important to highlight important points via this chapter prior next level of research works is done.



CHAPTER TWO

LITERATURE REVIEW

2.0 Introduction

This chapter reviews existing literatures related to the context of this research which are the Islamic Financial Industry (IFI) which covers technology adoption, Shariah decision making process, and main legal sources for the Islamic financial industry. The chapter also discusses the subject of this research which are information retrieval models, artificial intelligence (AI), machine learning (ML), and NLP. It ends with review on existing literatures on NLP.

2.1 Islamic Financial Industry (IFI)

The emergence of the Islamic financial industry since 1960s have brought to the practiced of dual banking system in the world. The system has been recognized not only in the Muslim countries but also by the non-Muslim countries. This recognition is something that must be proud of by the Muslim as the world have respected the Islamic law (Shariah) as the main law for the Islamic financial businesses. This is because the aim of the Shariah is to safeguards all aspects of human being and this is the missing part of the conventional financial system. The objective of the Shariah (*Maqasid Shariah*) which comprised of five elements (religion/belief, lineage, wealth, intellect, and life) became the main aspects whenever the Islamic financial products is offered in the market.

Accordingly, with the strong support from the authority and Shariah advisory committee both at local and international level such as Shariah Advisory Council of Central

Bank of Malaysia and the International Islamic Fiqh Academy which comprised of 57 member states of the Organization of Islamic Cooperation (OIC), any issues arise pertaining to the industry is discussed and solved at various levels. Now, the industry have been developed, remain competitive with the conventional counterparts, and evolves with the financial technology. As an update of the Malaysian Islamic financial industry, the central bank has granted five digital banks licenses recently on April 2022 which are;

Licensed under the Financial Services Act 2013 (FSA):

- i. a consortium of Boost Holdings Sdn. Bhd. and RHB Bank Berhad;
- ii. a consortium led by GXS Bank Pte. Ltd. and Kuok Brothers Sdn. Bhd; and
- iii. a consortium led by Sea Limited and YTL Digital Capital Sdn Bhd.

Licensed under the Islamic Financial Services Act 2013 (IFSA):

- iv. a consortium of AEON Financial Service Co., Ltd., AEON Credit Service (M) Berhad and MoneyLion Inc.; and
- v. a consortium led by KAF Investment Bank Sdn. Bhd.

Three out of the five consortiums are majority-owned by Malaysians namely Boost Holdings and RHB Bank Berhad, Sea Limited and YTL Digital Capital Sdn. Bhd. and KAF Investment Bank Sdn. Bhd.

2.1.1 Financial Technology (Fintech)

The word Fintech is the integration between “finance” and “technology” have captured the attention from the public since a decade ago (Liudmila, 2016). Anjan (2020) defined it as the use of technology to provide new and improved financial services. Discussing on Fintech, Luidmila et al. (2016) viewed it as a very broad phenomenon as it changed day by day

according to the need of the society. Some scholar opined that Fintech are innovation of payment systems including cryptocurrencies, credit markets and insurance (Anjan, 2020). The number of investments into Fintech companies and start-ups has increased tremendously within one year from USD 4.05 billion in 2013 to USD 12.2 billion in 2014 (Skan et al., 2015) and in Malaysia due to the pandemic Covid19, the year 2021 became a transformative year for the financial services where there was a shift in payment behaviour from a traditional to a digital payment system (Malaysia Fintech Report 2021). Digital payments and e-wallets are leading the Malaysian fintech space in the 2021 with 20% and 15% market shares accordingly (Malaysia Fintech Report 2021). Compared to the period prior the pandemic Covid19, the number of Fintech players has increased 40% which was 233 in the 2021 and 166 in the 2018. According to Mr. Salim, a Chief Executive Officer (CEO) of BigPay (the leading fintech company), by 2040, it's projected that one billion relatively young and digitally savvy people will demand financial services across Southeast Asia and Malaysia become the founder for the digital economy from the overhaul of domestic payments with PayNet, to digital identity frameworks, open-banking and data sharing standards (Malaysia Fintech Report, 2021).

Accordingly, there is a positive feedback and initiatives on the digital economy around the globe. In Malaysia, the central bank has stated vision to capitalise on digitalisation via its Final Sector Blueprint 2022-2026. To ensure the financial institutions to remain competitive, the central bank have confirmed its support towards technology and data-driven innovation. For example, in November 2021, the central bank and Payments Network Malaysia Sdn. Bhd. (PayNet) have collaborated with the international players on proof of cocept (POC) to enhance cost and speed of cross-border transactions. In addition to the digital bank licenses, the new

initiative by the central bank is issuing Digital Insurance and Takaful Framework to facilitate the licensing of digital insurers and Takaful operators.

2.1.2 Fintech and Banking Business

The current fintech environment has opened unprecedented opportunities for banking businesses and their customers. It is undeniable that the financial services have now become faster and easy to use where money can be transferred within few seconds and financing products can be approved within hour. Technology has replaced human roles in providing financial services. Scholars have agreed that fintech somehow has indirectly bring negative effect to the traditional banking system (Petralia et al., 2019) and (Berg et al., 2019).

In Malaysia, the government through the central bank have supported the emerging of technology in the Malaysian financial system. Among the initiative of the bank is to capitalise on digitalisation which stated in Malaysia Financial Sector Blueprint 2022-2026. The bank has provided few technology infrastructures as a backbone of the digital economy which includes real time payment systems and few guidelines on fintech. After a successful establishment of digital banks, the central bank has issued a discussion paper on the Licensing Framework for Digital Insurers and Takaful operators in February 2022. Other standards or guidelines that was set by the bank are policy on Risk Management in Technology (2020), Financial Technology Regulatory Sandbox Framework (2016), and Minimum Guideline on the Provision of Internet Banking Services by Licenced Baking Institutions (2014).

Overall, the technology has also open door to new businesses related to financial services (fintech business) in which this has become a challenge to the existing financial

institutions. However, it is observed that the Malaysian financial market is healthy enough to welcome newcomers where they have complimented each other and even become a strategic partner. In this context, the Malaysian government have been supportive and among the initiative given to the small medium enterprises fintech companies are tax angle incentive (granted to angle investors in fintech start-ups), income tax exemption by Malaysian Industrial Development Authority (MIDA), partial corporate tax exemption for entities in the Malaysian Digital Hub under MDEC, Malaysia Tech Entrepreneur Program under MDEC, Multimedia Super Corridor (MSC) Malaysia status recognition for ICT, and others.

2.1.3 Shariah Decision Making Process in IFI

In Malaysia, the highest authority that is responsible to advise IFI on Shariah matters at the Central Bank level is known as Shariah Advisory Council (SAC)¹ whereby at the respective Islamic bank level, it is called Shariah Committee (SC)² or some institutions called it Shariah Advisory Board (SAB). At the Central Bank level, Section 52(1) of the Central Bank of Malaysia Act 2009 have indicated that the function of SAC are; (a) *to ascertain the Islamic law on any financial matter and issue a ruling upon reference made to it in accordance with this Part; (b) to advise the Bank on any Shariah issue relating to Islamic financial business, the activities or transactions of the Bank; (c) to provide advice to any Islamic financial institution or any other person as may be provided under any written law; and (d) such other functions as may be determined by the Central Bank* (page 4). Meanwhile, establishment of

¹ Based on IFSA 2013, SAC means the Shariah Advisory Council on Islamic finance established under section 51 of the Central Bank of Malaysia Act 2009

² Based on IFSA 2013, SC means the Shariah committee of an institution established pursuant to section 30

SC by Islamic banks for the purposes of advising the business, affairs, and activities to comply with the Shariah is mandated in Section 30 (1) of IFSA 2013.

According to Abdullah et al. (2014), there are three stages of product approval generally practiced in the financial industry which are pre-product approval, during product approval, and post product approval process. Below is summary of the process.

Table 2. 1: Product Approval Process

Product Approval Process			
	Pre-approval	During-approval	Post-approval
Tasks	Product structuring/design	Checking if there is any Shariah issues appear during product development process	Monitoring product implementation to ensure compliance with the Shariah principles
	Shariah research	Advise and decision must be made in a comprehensive manner	Identifying the area of potential Shariah non-compliance risks
	Product vetting	Explain all the process involved, documents used and other necessary information	Proposing the relevant actions to management
	Agreement vetting	All new products should be approved by Shariah committee and backed by fiqh literature, evidence, and reasoning	Conduct internal Shariah review and Shariah governance reporting
	Compliance checking	Rigorous deliberation process among SC on the legal contract and other documents	Conduct Shariah audit to ensure the business is fully comply with the Shariah. A periodical assessment to provide independent assessment and objective assurance designed to add

			value and improve degree of compliance of the IFI businesses.
	A review of concept, structure, contract, term sheet, documentations, policies, and procedures, pamphlets, brochures, and advertising	Professional judgement among SC to reach a consensus in arriving decision Consistent opinions of SC in serving different institutions	

Overall, the most important issue that concerned by the central bank is compliancy of the Islamic banking businesses towards Shariah. The establishment of advisory committee whether at the central or respective bank level would be very crucial to ensure that the Islamic banking is practiced accordingly to meet the objectives of the Shariah which is for a betterment of ummah (society). In this context, the central bank has come committed to confirm the roles of all related stakeholders in the IFI including board of director, senior management, SC, as well as Shariah secretariat and control Functions (Shariah risk management, audit, review) via Shariah Governance Policy Document (SGPD) in 2019.

2.2 Information Retrieval

Information Retrieval (IR) or text information retrieval is basically denoted the ‘information’. It is a branch of computer science that involves with activity of processing as well as searching of unstructured documents (Nadkarni, 2002). It caters information needed by users pertaining to their query (Sharma & Kumar, 2020). It overlaps with other computer science fields like database technology and natural language processing (NLP). The objective of IR basically to

assist text searching activity in order to obtain fast and accurate information based on key words. Nowadays people rely on IR system through index-based search engines like a Google or others to find the information that they need even though it does not give specific answers to the question. This limitation still exists until the current decade where a recent study conducted by Sharma and Kumar (2020) have mentioned that there are two major issues for IR which first, elusiveness in specifying user's information need through query and second, limited number of methods to perform partial match between documents and query. In this context, the authors have suggested intelligent information retrieval model which is called Bayesian Rough Set and the model is found to perform well.

Accordingly, Intelligent Information Retrieval (IIR) is introduced to compliment the limitation of IR where it is known as 'agent' information. The idea of 'intelligent' information retrieval was first introduced in the late 1970s and have become popular in the early 1990s (Belkin, 1996). Under one single umbrella, IIR works as information gathering agent, extracts both unstructured and structured documents, and provides index for information searching activities (Gupta et al., 2012). IIR has been defined variously by different scholars, and a consistent interpretation is it reflects the machine program doing something for the user, or taking over some functions that previously had to be performed by humans (either user or intermediary), retrieves and forwards information to the user without any other interaction. (Belkin, 1996). In this context, the usage of artificial intelligence (AI) in IIR would enhance the efficiency of information retrieval activities (Xie et al., 2003). It basically integrates various AI technologies including scheduling, planning, text processing, information extraction and interpretation (Gupta et al., 2012). Among advantages of IIR are it supports human decision-

making process, reducing the manual workload, and time spent to fulfil tasks (Gupta et al., 2012; Lilis & Scanlon, 2016).

2.3 IR Retrieval Process

An Information Retrieval System (IRS) would aid to find document that is needed by user (query) as well as find similarities among the documents extracted (Al-Baatla et al., 2019). Three main components of IRS are query subsystem, document database, and matching function (Al-Baatla et al, 2019; Pathak et al., 2000). All documents are stored in database, and whenever the users write their queries, all documents needed will be presented to them. On top of that, IRS matching function is responsible to match all documents in the database with users' query. The processes are presented in Figure 2.1 below (squared boxes represent data and rounded boxes represent processes).

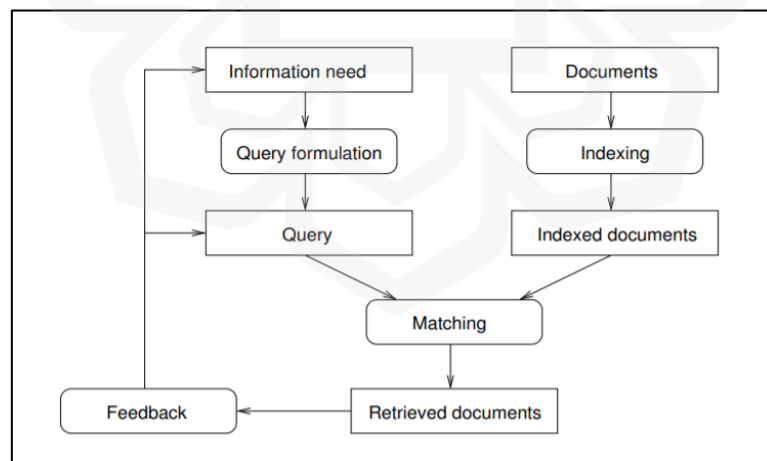


Figure 2.1: IR Process (source: Goeker & Davis, 2009)

2.3.1 Dimensions of Similarity

Documents might be declared as similar based on certain criteria, for example when they cover the same topic, use a common set of words or are written in the same style. In IR, the dimension of similarity defines the understanding of overall text similarity. Three dimensions of similarity that was known in the literatures are lexical, structural, and semantic similarity (Lin, 1998). Not all cases suitable to be compared in terms of similarity as they are subjective. Document similarity is usually measured through scalar score, which is within an interval of zero to one. The highest degree of similarity is measured as one. When two objects are dissimilar, the degree of similarity is zero (Lin, 1998).

2.3.1.1 Lexical Similarity

Lexical similarity is calculated based on a simple word-to-word comparison (Lovins, 1968). Methods like stemming or stop word removal may increase the result of lexical similarity. It relies on the words of document text. A total overlap between words would result in a lexical similarity of 1, whereas 0 means both documents share no words or not similar (Fox, 1989).

2.3.1.2 Structural Similarity

This type of similarity focuses on the vocabulary where it explains the theoretical composition of two documents (Shin & Doermann, 2006). It basically covers graphical components (text layout and design) as well as composition of text segments including paragraphs or sentences, citations format, and hyperlinks (Buttler, 2004). This similarity measures mostly adopted for semi-structured document such as XML or HTML. A common and expensive in computing time, structural similarity is performed to calculate the minimum cost edit distance between two documents which measures the number of actions that are needed to change a document, so it is equal to another document (Shin & Doermann, 2006).

2.3.1.3 Semantic Similarity

Semantic similarity is more focused on similarity of topics of two documents that have similar meaning. An appropriate semantic similarity measures is important for many IR systems, given that the users need information based on semantic meaning instead of vocabulary or structure of a document. Hence, measuring topical relatedness is not an easy task which need a proper algorithm. Therefore, lexical or structural similarity is often used to estimate semantic similarity (Buttler, 2004).

2.4 IR Model

There are no specific and best model that suit all scenario of IR process. Goeker and Davis (2009) have highlighted three IR models as summarised below.

2.4.1 Exact Match Model

2.4.1.1 Boolean Model

The Boolean model is one of the earliest and simplest information retrieval models (Bhatnagar et al. (2012). It is based on George Boole's mathematical logic where queries and their equivalent texts is combined to create a new set of documents (Goeker & Davis, 2009). In other words, a document is represented as a set of keywords where queries are presented by Boolean expressions of keywords, connected by AND, OR, and NOT, including the use of brackets to indicate scope. The output would be either the document is relevant or not. No partial matches or ranking in this IR model. It is said as the weakest model because it fails in the partial match (Al-Baatla et al., 2019). Its only popular at earlier time because of simple and easy to understand. Few disadvantages of this model including it does not offer a ranking of retrieved documents, very rigid (AND means all; OR means any), difficult to response complex user

queries and too complex for general web users (Goeker & Davis, 2009; Al-Baatla et al., 2019).

Among advantages of this model are it gives users a power to control the system and immediately clear why a document has been retrieved given a query.

2.4.1.2 Region Model

Region model is extension of the Boolean model that explains the subjective parts of document which is called segments or regions. Any sequence of consecutive words is called a region and identified by a start and an end position. This model does not limit to retrieving documents as it would depend on the queries. The Boolean operators AND, OR and NOT might be defined on sets of regions in a straightforward way as set intersection, set union and set complement. Two additional operators introduced by this model are containing and contained by (Goeker & Davis, 2009).

Overall, the Boolean model is grounded in mathematics and its instinctive use of document sets for information retrieval. The major disadvantage of the Boolean and the region models is their failure to rank documents. In IR, document ranking is the most concerned by user and that is the reasons the next models would cater the weaknesses on the Boolean and region models. The following sections will discuss few more extension models of ranked retrieval.

2.4.2 Vector Space Model (VSM)

Luhn (1957) is the first scholar that propose a statistical method to search information (Goeker & Davis, 2009). For him, to discover a document collection, a similar document should be prepared to the documents searched. Then, the similarity between the document prepared with

the one in the collection is used to rank the results. The similarity criterion is elucidated by Luhn (1957) as follows:

The more two representations agreed in given elements and their distribution, the higher would be the probability of their representing similar information. (Goeker & Davis, 2009)

In 1980s, Gerard Salton and friends have adopted Luhn (1957)s' similarity criterion and proposed VSM where query is termed as vector embedded in high dimensional Euclidean space and each item is assigned a distinct dimension called item (Goeker & Davis, 2009). It is necessary to convert a document to a vector or mapping the documents (de Jong and Frederiksen, 2015) in the vector space, then use the vector calculation to compute the distance between the vectors. The objective of the VSM is to simplify and calculate the complex relationship of words between two independent documents that is expressed in a vector (Indriyanto & Sumitra, 2019). Generally, the key to establish VSM is to determine the dimension and it corresponds to one of the keywords in document, where the dimensions of the model would relate to the total number of keywords in the entire documents.

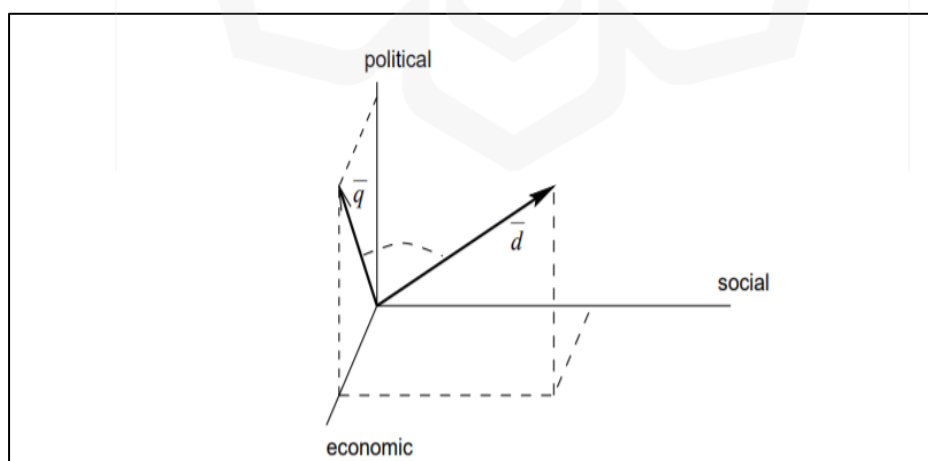


Figure 2.2: Sample of a query and document representation in the vector space model

As presented in Figure 2.2 above, the similarity measure is referred to the cosine of the angle that separates the two vectors 'd' and 'q'. The cosine of an angle is 0 if the vectors are orthogonal in the multidimensional space and 1 if the angle is 0 degrees. Cosine would measure the level of closeness of the two vectors. The cosine formula is given by:

$$\text{score}(\vec{d}, \vec{q}) = \frac{\sum_{k=1}^m d_k \cdot q_k}{\sqrt{\sum_{k=1}^m (d_k)^2} \cdot \sqrt{\sum_{k=1}^m (q_k)^2}}$$

Based on Luhn (1957) similarity belief, the first step is to count the number of elements that the query (prepared document) and collection documents shared. If the document vector represented by vector $d = (d_1, d_2, \dots, d_m)$ and query represented by vector $q = (q_1, q_2, \dots, q_m)$, with similar component d_k ($1 < k < m$), so the similarity measure is;

$$\text{score}(\vec{d}, \vec{q}) = \sum_{k=1}^m d_k \cdot q_k$$

Among advantages of VSM are it is a simple, practical, and provides automatic text categorisation, clustering, partial, and true matching (Goeker & Davis, 2009; Indriyanto & Sumitra, 2019; Kadowaki, & Kishida, 2020). Few weaknesses of this model are it does not concern on syntactic and semantic information like phrase structure, word order, proximity information, and word sense. VSM has been adopted by past researchers from various area including the prominent area from natural language processing (Dongmei Han et al., 2018), education (Indriyanto & Sumitra, 2019; Foltz & Landauer, 1999; Jiffriya et al., 2014), medical (Thakur et al., 2019), and others.

2.4.3 Probabilistic models

This model is based on probability theory that assign a weighting for possible terms through various experiments. Probability model will rank documents in decreasing order of probability of relevance to the information need: $P(R|q,d_i)$ based on similarity coefficient between a query and a document. This model was implemented for small-scale search tasks.

$$d_j = (t_{1,j}, t_{2,j}, \dots, t_{i,j}), t_{i,j} \in \{0,1\}$$

terms occurrences are boolean (not counts)

query q is represented similarly

R is the set of relevant documents, $\sim R$ is the set of irrelevant documents

$P(R | d_j)$ is probability that d_j is relevant, $P(\sim R | d_j)$ irrelevant

Table 3: Comparison of IR models (Thakur et al., 2019)

IR Criteria / IR Retrieval Model	Boolean Model	Boolean Model variants	Extended / Soft Boolean Model	Vector Space Model	Probabilistic Model
Information Retrieval Model	Deterministic	Deterministic	Deterministic	Deterministic	Probabilistic
Indexing	Complete items	Complete items	Complete items	Complete items	Derived from the content
Matching Retrieval	Exact Match	Exact Match	Exact Match	Partial or Best Match	Partial or Best Match
Query type	Structural	Structural	Structural	Structural	Natural Language
Result criteria	Any Match	Any Match	Any Match	Relevance	Relevance
Result ordering	Arbitrary	Arbitrary	Ranked	Ranked	Ranked

Table 3: Strengths and Weaknesses of the Information Retrieval models

Model	Strengths	Weaknesses
Boolean Model And Boolean Model variants	The retrieved documents can be either limited or voluminous and also relevant or irrelevant.	The size of the resultset is unpredictable where it can be either too many retrieved documents or none. It considers all the retrieved documents in the resultant posting list. Since all terms are weighted equally, the retrieved documents will not be ranked. Hence, all documents are considered "equally worthy." Documents that "don't fairly matched" the query may be beneficial also. There is no provision for partial matches
Extended / Soft Boolean Model	It is a simple model based on Linear Algebra and Term weights. The term weights are not binary. This model allows partial document matching.	Formulating useful extended Boolean model requires more thought and expertise in the query domain.
Vector Space Model	It is a Term weight model based on a Geometric similarity measure. It uses the Dot product of the query and document vector and can even allow partial matching.	Lengthy documents have little similarity scores. Precise match of the query keywords in the document terms may result in false positive results due to substring match.
Probabilistic Model	Given a query, the model ranks documents by the probability of relevance.	Independent assumption and Parameter estimation are the two crucial principal concern issues. It is hard to estimate parameters, i.e. need to estimate relevance, without a proper training dataset. In the composite terms, the presence of one term increases the likelihood of the presence of the other even though sometimes it is not realistic.

2.5 Literatures on Natural Language Processing (NLP)

Natural language processing (NLP) is a subset of artificial intelligence (AI) that helps computers understand, translate, and manipulate human language. NLP basically implement tasks such as simply language, break it into pieces, and explain the relationships among those

pieces to generate meaningful content. The combination of NLP, machine learning and human experts would modern problems is tackled.

The applications of NLP are diverse and ideal for nearly any situation involving the need to rapidly analyze unstructured text. There were many research from health and sciences have utilized NLP to trace diseases from any symptoms. Patterns of symptoms and root cause analysis would be nearly impossible for a human to detect by observing every individual record. However, an AI system can work around the clock to analyze the test results, patient reports, listed symptoms and more. NLP has proven powerful in its application to predict and alert hospitals when data in electronic health records indicates the presence of sepsis.

NLP have been used to other domains such as to analyze legal case files, social media feeds, call center logs, research documents, warranty claims and others. The majority of data held by organizations is in the form of unstructured text. NLP would basically perform linguistic analysis and converts text into form representations for text processing and understanding. This includes methods such as tokenization, removing stop words, stemming, and others. The output from NLP then is utilised by machine learning (ML) for data mining to automate the production of main input of the texts for descriptive analytics. In this stage, human input such as knowledge, experience, and expertise are important to further explain the machine output.

NLP is a powerful tool to enhance our living and working environment. The combination between human and technology in terms of information sharing would be more meaningful especially for the unstructured information where it could bring progress to areas

that have been slow or difficult to solve previously. It may assist human to deliver better input, experiences, improve quality of care and enrich human engagement.

2.5.1 Existing Works on NLP

In Islamic finance industry, Haider et al. (2020) have suggested the usage of AI in NLP based on Islamic FinTech Model that combines Zakat and Qardh-Al-Hasan (benevolent loan) to minimize the negative impact of COVID 19 on individuals and SMEs. The authors were in opinion that the Islamic finance has big potential to face any kind of situation/pandemic especially the combination of Zakat and Qardh-Al-Hasan. In addition, research done by Ali et al. (2019) and others (Laldin & Furqani, 2019; Zubaidi & Abdullah, 2017; Evans, 2015) have explored the potential of Fintech which including AI, smart contract, blockchain, and others on the Islamic banking and finance industry in various Asian countries and their findings indicate that Fintech would benefit the industry greatly to be at par with the conventional counterparts. Hence, a lot of effort need to be done to implement Fintech in the Islamic financial industry so that it would bring significant impact to the society and practiced with the Shariah spirit.

Figure 2.3 presents the increase trend of literatures on NLP based on Scopus database in all areas of research which including social science, arts, engineering, computer sciences, business, management, economics, and others. Since, the 2016, the number of literatures remain increase until year 2021 (2045) and is expected to further increase until end of this year. Since 1976, a total of ten thousand of articles have been published in Scopus database and the author that have contributed a lot in this domain is Liu, H with 77 articles. It is followed by Xu, H and Friedman, C. Please refer to next Figure 2.4.

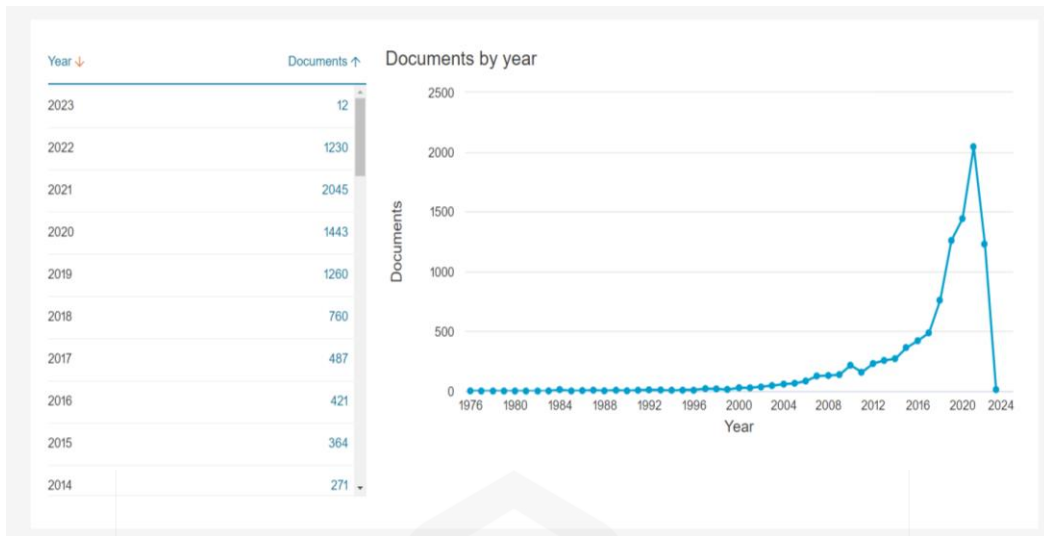


Figure 2.3: Number of literatures on NLP by year

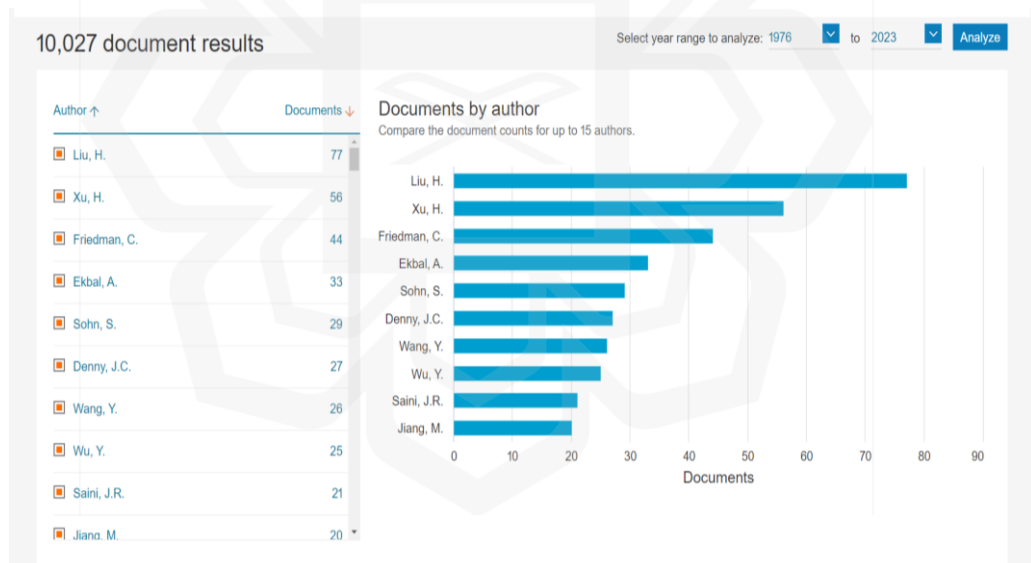


Figure 2.4: Total number of NLP literatures by author

Accordingly, as presented in Figure 2.5 below, United States has been the dominant in discussing NLP (2642 articles), followed by China (1542 articles), India (1356), and United Kingdom (573 articles). In is undeniable that these countries have been the forefront of technology adoption. It is also not a surprise where NLP is mainly conducted by researchers

from computer science (37.3 %) and engineering (14.3%). Other areas that have adopted NLP are medicine, social science, arts, physics, business, management, and accounting. Please refer to Figure 2.6.

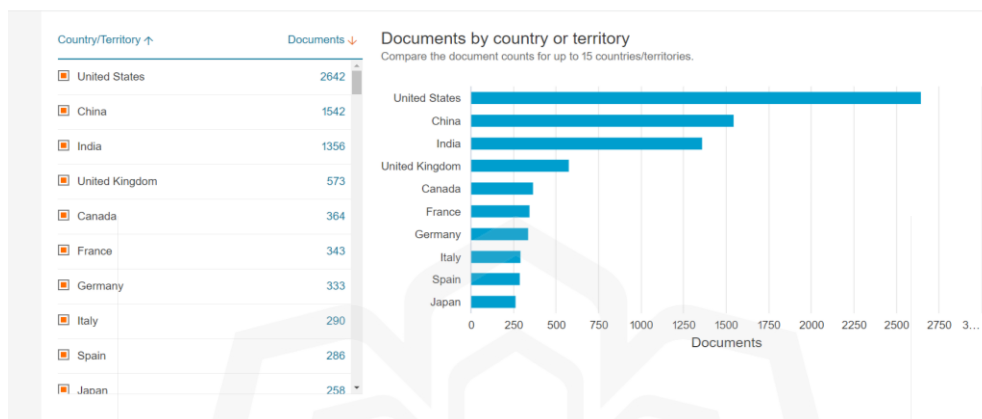


Figure 2.5: Countries that have contributed to NLP literatures

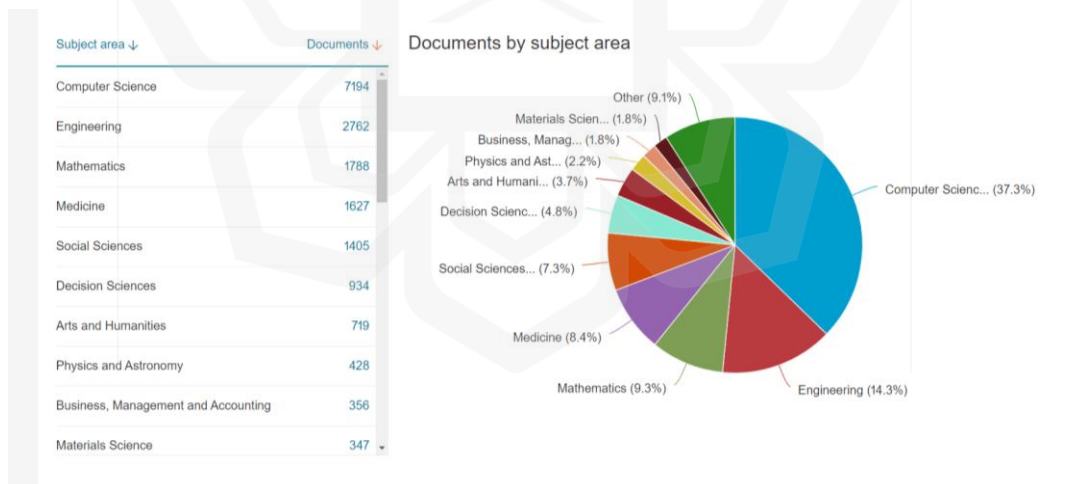


Figure 2.6: Literatures on NLP by subject area

In addition, out of ten thousand literatures on NLP from various area, this research have limited the search of subject area to only social sciences, business,

management, and accounting. As in Figure 2.7, there is 1712 articles found in these areas. Bhattacharyya, P., Ekbal, A., and Korhonen, A. has contributed equally to the literatures.

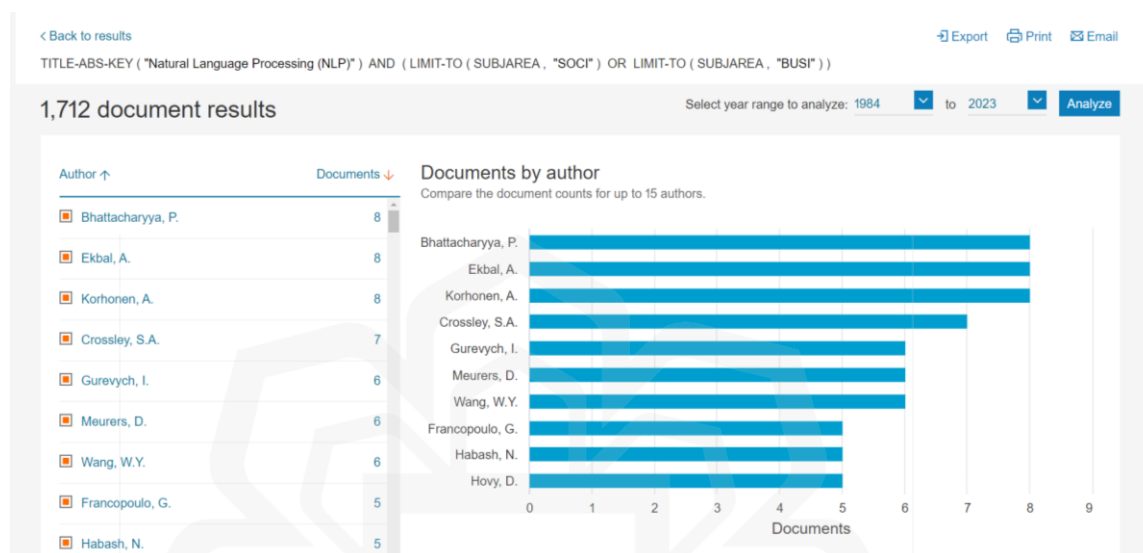


Figure 2.7: NLP literatures on social sciences, business, management, and accounting areas

This research further defined the search to ‘economics, finance, business, management, and accounting’ subject area. The results indicate that, there are 178 articles with increasing trend and expected to increase until 2022. Please refer to Figure 2.8. The main contributors are Coffas, Delcea, and Melumad where each of them wrote 3 articles on NLP between year 2019-2020. Most of the authors have adopted NLP on social media research either in terms of sentiment analysis based on image evaluation, customers’ product opinions, social media users’ emotions, or on consumers vocabularies. The most cited article was by Melumad (2019) which on effect of social media content by usage of smartphone. Based on literatures findings, there is still limited literatures that have utilised NLP for banking and finance.

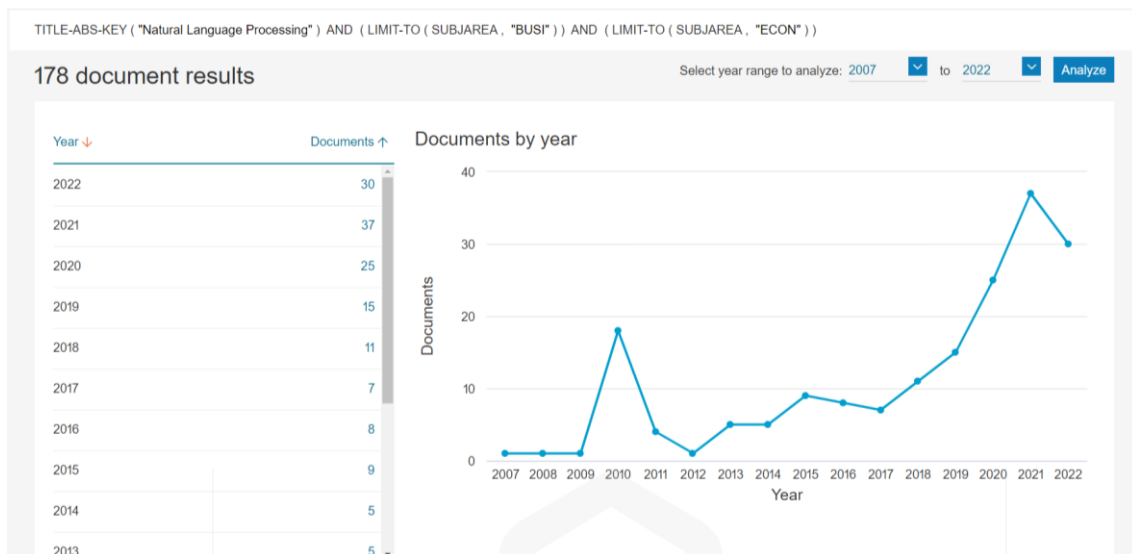


Figure 2.8: NLP literatures from business, economics, and finance area

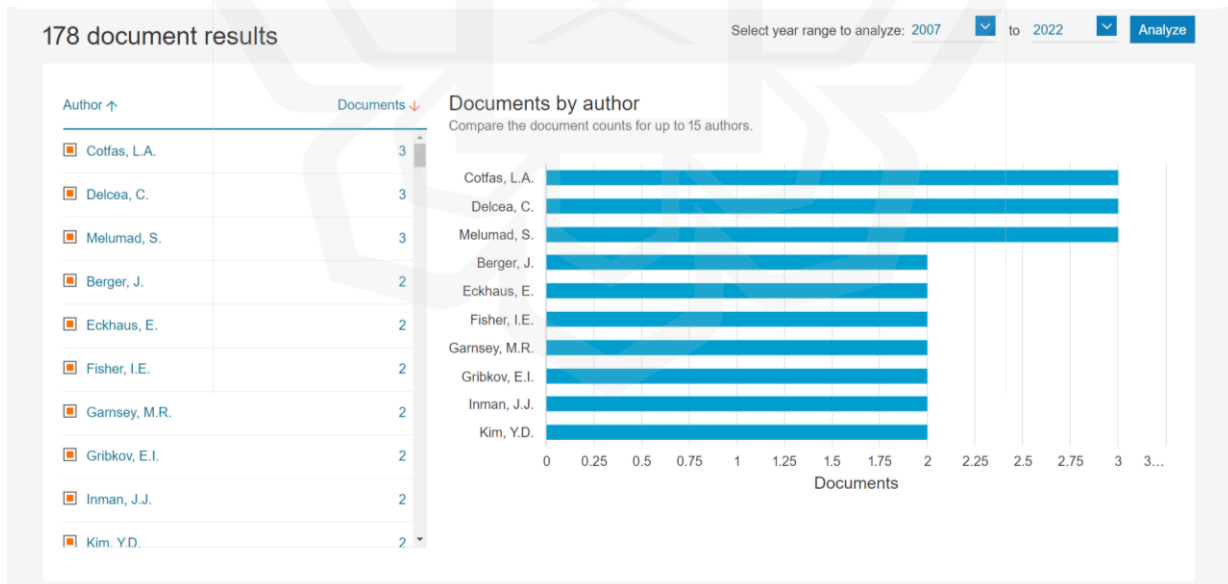
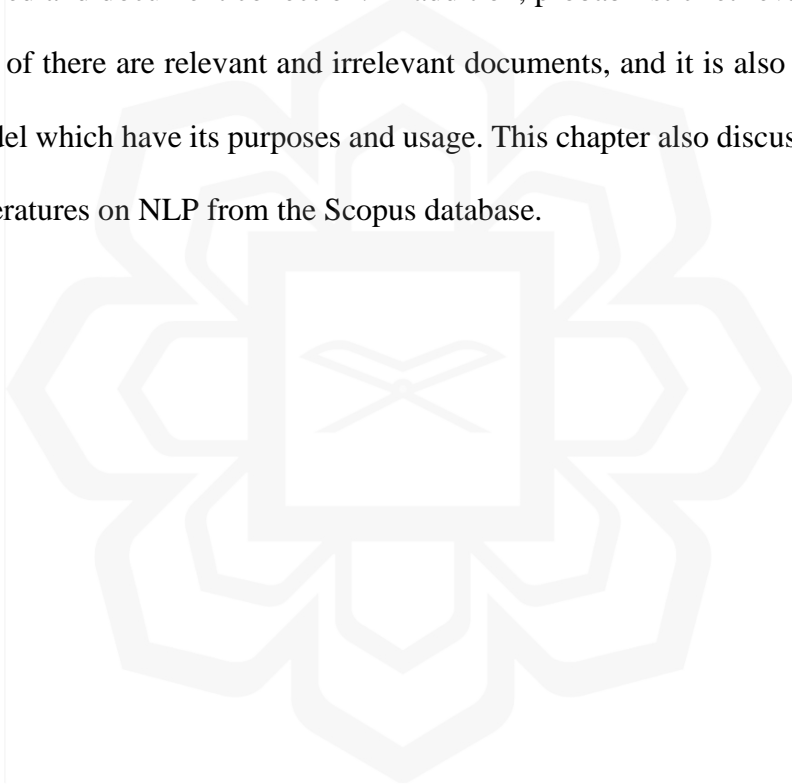


Figure 2. 9: Main contributors on NLP literatures

2.6 Chapter Summary

First part of this chapter reviews the existing literatures on financial industry mainly related to technology and Shariah decision making process. This chapter then discusses various models in IR. There is no such thing as a dominating model or theory of IR. In IR, some models work for some applications, whereas others work for other applications. For example, the region models is designed to search semi-structured data and the VSM is more suit for similarity search of prepared and document collection. In addition, probabilistic retrieval model is more suitable in case of there are relevant and irrelevant documents, and it is also same with other types of IR model which have its purposes and usage. This chapter also discusses what is NLP and existing literatures on NLP from the Scopus database.



CHAPTER THREE

RESEARCH METHODOLOGY

3.0 Introduction

This chapter basically highlights the methodology adopted in this research to achieve the research objective. As mentioned earlier, based on literatures, Vector Space Model (VSM) is found to suit with the research objective which is to find pattern matching between two types of documents which is the one proposed by the bank and the guidelines imposed by the regulator (Central bank Malaysia). This chapter first discusses preprocessing procedures using Natural Language Tool Kit (NLTK) and followed by similarity analysis based on Python language.

3.1 Research Design

Research design comprises of the overall strategy of the research to ensure an effective way to address the research problem; it consist of theoretical underpinning, data collection, measurement and data analysis (Hussey & Hussey, 1997). According to De Vaus (2001), research design is not only about research strategy, but it also ensures that the outcome obtained from the research is reliable and unambiguous to reflect the research questions. Research can be designed qualitatively, quantitatively, or mixed methodology. This research is designed based on mixed methodology approach. It involves both quantitative as well as qualitative method for data collection and analysis. Qualitative approach is first done by collecting bank proposal and guidelines from the Central Bank of Malaysia for document similarity algorithms building. The second stage is quantitative approach where data is collected through online

structured survey to industry practitioners to observe their view on the Shariah screening application proposal. Below is framework of research design for this research.

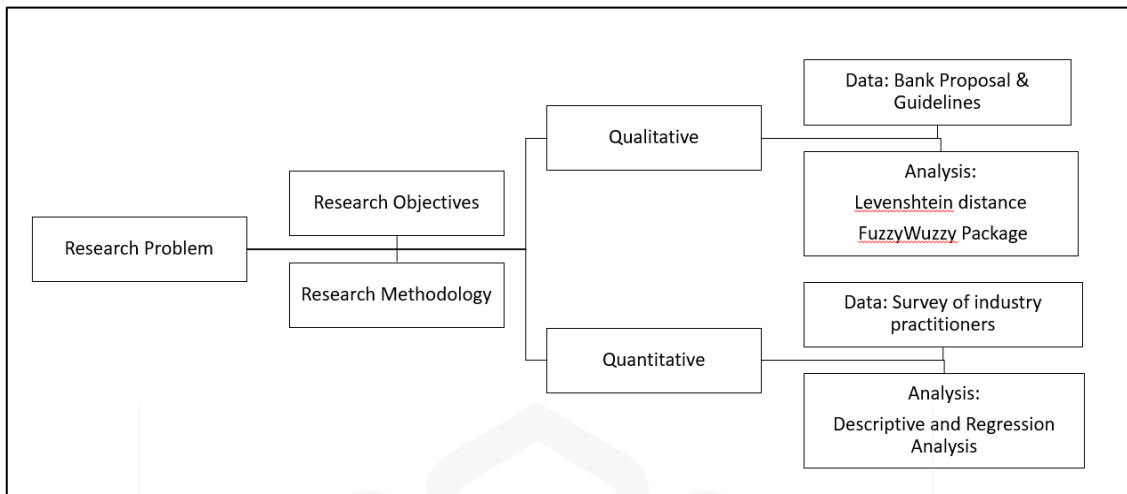


Figure 3.1: Research Design

3.2 Proposed Algorithm

The proposed model in this research is based on VSM for document matching and similarity measures. Product proposal will become a query and policy documents by the central bank would be a corpus or database for document matching. Both the query and corpus went through preprocessing stage prior similarity analysis. One set of queries with two set of corpuses is tested in this research to compare similarity value. For document similarity checking, this research adopts Fuzzy String matching consists of Levenshtein Distance and FuzzyWuzzy Package method by Python programming language. It is a technique of finding strings that match with a given string partially and not exactly. When a user misspells a word or enters a word partially, fuzzy string matching helps in finding the right word. The algorithm behind fuzzy string matching does not simply look at the equivalency of two strings but rather quantifies how close two strings are to one another. This is usually done using a distance metric known as ‘edit distance’. This determines the closeness of two strings by identifying the minimum alterations needed to be done to convert one string into another. There are different

types of edit distances that can be used like Levenshtein distance, Hamming distance, Jaro distance, and other. Hence, this research adopts Levenshtein distance and FuzzyWuzzy Package to check similarity of the document. Figure 3.2 below presents the proposed algorithm for document screening.

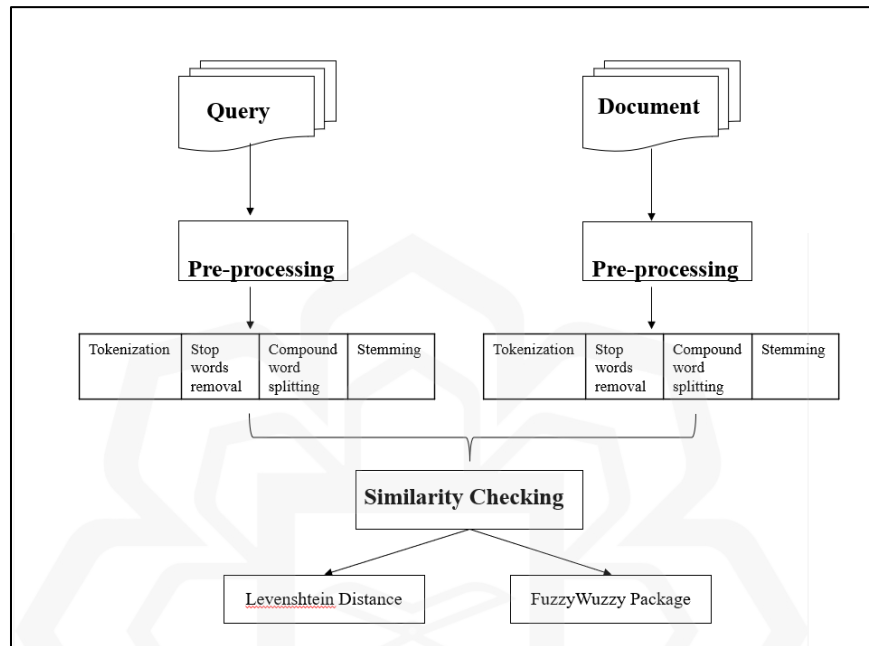


Figure 3. 2: Proposed algorithm for document screening

3.3 Data Preprocessing

Two sets of documents (product proposal and BNM guidelines) were selected in this research and went through preprocessing stage to avoid any irrelevant information or accuracy of the results later. Steps in data pre-processing are includes tokenization, stop-word removal and stemming (Manning et al., 2008). The documents first gone through tokenization which the whole documents were transferred into words using white spaces. It was followed by punctuations and stop words removal like comma, semicolon, ‘and’, ‘is’, ‘or’, and others. Next, any compound words were split, and stemming were done using Porter stemming program. In

this stage, words were converted into their stems such as ‘structuring’ or ‘structures’ into ‘structure’ to determine domain vocabularies and to reduce redundancy as most of the time the word stem and their derived words mean the same. Once, this preprocessing stage was completed, the similarity analysis was done.

3.4 Measure of Similarity

Once the queries and corpus have been screened and ready for further analysis, the vital process is checking for documents similarities. Similarity analysis is a necessary stage in most of IR and NLP tasks including for document clustering (Huang, 2008), plagiarism detection, text categorization (Bigi, 2003), and document screening. The success of IR models mostly depended on its similarity measures (Poletini, 2004). There were various measures of similarity as discussed in the past literatures whereby the differences among the measures were on its functionality; a similarity measure that is effective in addressing one measurement problem may not be effective in another (Thomson et al., 2015). This research adopted two methods of fuzzy string by Python which were Levenshtein Distance and Fuzzy Wuzzy Package.

3.4.1 Levenshtein Distance

Levenshtein Distance or known as Minimum Edit distance is a popular method used to measure the distance between two strings. It is computed by counting number of edits required to transform one string into another. The edits could be either addition of a new letter, removal of a letter, and substitution. As for example, the Levenshtein distance between “house” and “mouse” is 1 as only 1 edit is required to change ‘h’ into ‘m’. There could be multiple ways of transitioning from one word to another, but Levenshtein distance chooses the smallest possible path. The more similar between two strings the less distance between them, and vice versa. This method is commonly used in autocompletion or autocorrection application of text like

Google search or online dictionary. Four processes will involve to come out with distance values including creating the distances matrix, initializing the distances matrix, printing the distances matrix, and finally calculating distances between all prefixes (Gad, 2020).

Accordingly, this method is introduced by Vladimir Levenshtein in year 1965. This is Mathematical formula created by him to measure similarity of distance.

$$\text{lev}_{a,b}(i,j) = \begin{cases} \max(i,j) & \text{if } \min(i,j) = 0, \\ \min \begin{cases} \text{lev}_{a,b}(i-1,j) + 1 \\ \text{lev}_{a,b}(i,j-1) + 1 \\ \text{lev}_{a,b}(i-1,j-1) + 1_{(a \neq b)} \end{cases} & \text{otherwise.} \end{cases}$$

Figure 3. 3: Levenshtein Mathematical Formula

Below is sample of python codes conducted in this research to measure the Levenshtein Distance.

```
min_distance = 1
max_ratio = 0
max_ratio_label = 0
max_ratio_label_content = ""
for nums in StrOptions.keys():
    Distance = lev.distance(String1.lower(), StrOptions[nums].lower())
    Ratio = lev.ratio(String1.lower(), StrOptions[nums].lower())
    print("Distance:", Distance, "Ratio:", Ratio, "", f'"{nums}"', StrOptions[nums])

    if max_ratio < Ratio:
        max_ratio = Ratio
```

```

max_ratio_label = nums
max_ratio_label_content = StrOptions[nums]

print('\n')

print("The least distance is:", min_distance, "The greatest ratio is:",
max_ratio, "\nTawarruq ", max_ratio_label, max_ratio_label_content)

```

3.4.2 FuzzyWuzzy Package

Fuzzywuzzy is a python library that is used to calculate the differences between sequences and patterns of two strings. There are four elements that can be used to measure the similarity in FuzzyWuzzy which are;

No.	Ratio Similarity Checking	Purpose	Code
1	fuzz.ratio()	For exact matching, order of string does matter	<code>from fuzzywuzzy import fuzz fuzz.ratio('', '')</code>
2	fuzz.partial_ratio()	For partial matching, order of string does not matter	<code>fuzz.partial_ratio ('', '')</code>
3	fuzz.token_sort_ratio()	Order of the words does not matter. It also ignores punctuation, and follows the concept of tokenization where the strings are converted in to tokens and then sorted in alphabetical order for comparison analysis	<code>fuzz.token_sort_ra tio('', '')</code>
4	fuzz.token_set_ratio()	Performs a set operation that takes out the common tokens (the intersection) and then makes	<code>fuzz.token_set_rat io('', '')</code>

		fuzz.ratio(). Extra or same repeated words do not matter.	
--	--	---	--



Below is sample of python codes conducted in this research to measure similarity using FuzzyWuzzy Package.

```
Ratio = fuzz.ratio(Str1.lower(),Str2.lower())  
Partial_Ratio = fuzz.partial_ratio(Str1.lower(),Str2.lower())  
Token_Sort_Ratio = fuzz.token_sort_ratio(Str1,Str2)  
Token_Set_Ratio = fuzz.token_set_ratio(Str1,Str2)  
  
print(Ratio/100)  
  
print(Partial_Ratio/100)  
  
print(Token_Sort_Ratio/100)  
  
print(Token_Set_Ratio/100)
```

3.5 Survey among Industry Practitioners

A structured survey among bank officers and Shariah advisors is carried out to investigate their opinion on the proposed automatic Shariah screening application. Due to time constraint to develop a real application, the respondents were given a short brief on the proposal (Automatic Shariah Screening Platform) in the survey form (Figure 3.3). The survey is conducted online and a total of ten industry practitioners have involved in the survey.

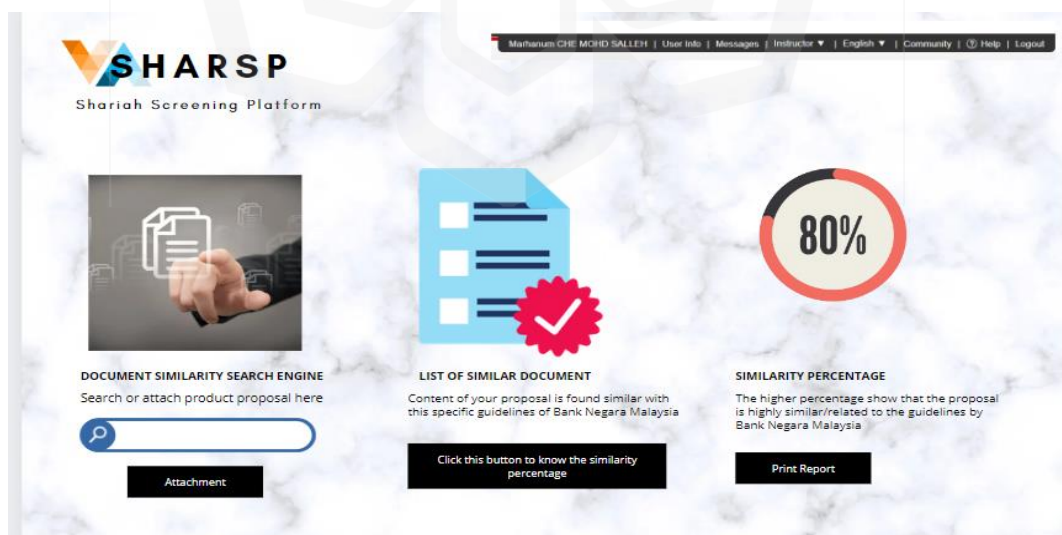


Figure 3.4: Sample of the Proposed Shariah Screening Platform

3.5.1 Technology Acceptance Model (TAM) and Technology Readiness Index (TRI)

Technology Acceptance Model (TAM) is adopted in this research to measure the acceptance towards the proposed automatic Shariah screening application among financial industry practitioners. TAM which was pioneered by Davis (1989) is focused more on analytical framework for predicting application of technological innovations. There are five key dimensions highlighted by TAM which are Perceived Usefulness (PU), Perceived Ease of Use (PEOU), Attitude (A), and Behavioural Intention (BI) (Davis, 1989). This research however adopts only three dimensions of TAM and adds one variable from Technology Readiness Index (TRI) by Parasuraman (2000) in the survey questions to suit with the survey objective which is to investigate practitioners' intention to adopt the proposed Shariah document screening platform. Five main questions which adopted from Kolade et al. (2022) and adapted in the survey are the following:

- 1) Do u think this app would be useful to check bank document/proposal for Shariah compliancy?
- 2) To what extent do you think this application are easy to use?
- 3) To what extent you are ready to adopt an automatic Shariah document screening platform?
- 4) Do you have an intention to use this application for Shariah screening?
- 5) Do you currently use any of technology/application to screen product proposal for Shariah compliancy?

3.5.2 Survey instrument

The survey instrument consists of open-ended questions. Part A is the respondents' background. Part B captured the feedback of the industry practitioners on the adoption of the proposed automatic Shariah document screening platform. Accordingly, this research adopts

survey instrument by Abu-Dalboah (2013) to measure Perceived Usefulness and Perceived Ease of Use, Technology Readiness (Parasuraman, 2000), and Intention to Use (ITU) by DuPree (2015). Below are the survey questions with 5-point Likert scales constructed in this research.

No.	Constructs	Operational Definitions	Measurement Items
1.	Perceived Usefulness (PU)	Feeling that industry practitioners assume in enhancing proposal preparation process by using automatic Shariah document screening platform	The Shariah document screening platform would assist the Shariah department to screen any Shariah non-compliance elements in the proposal.
			The Shariah document screening platform would shorten the time to prepare product proposal.
			The Shariah document screening platform would ensure the proposal is accurate according to the BNM guidelines/policy.
			The Shariah document screening platform would ensure the proposal having less mistake according to the BNM guidelines/policy.
2.	Perceived Ease of Use (PEOU)	Refers to level of easiness that the practitioners assume when using Shariah document screening platform.	Learning to utilise this platform would be ease for me.
			I would find it easy to use this platform.
			The platform seems easy and understandable.
			I would find the platform to be flexible and user friendly.
			It would be easy for me to become skillful at using this platform for Shariah document screening.
3.	Technology Readiness (TR)	Measures a person's technological preparation for new technology adoption	Optimism - Technology makes me more productive in doing my job - New technologies contribute to better quality of life - Technology gives me more freedom of mobility
			Innovativeness - In general, I am among the first in my circle of friends to acquire new technology when it appears

			<ul style="list-style-type: none"> - I can usually figure out new high-tech products and services without help from others - I keep up with the latest technological developments in my area of interest
			<p>Insecurity</p> <ul style="list-style-type: none"> - Technical support lines are not helpful because they don't explain things in terms I understand - Sometimes, I think that technology systems are not designed for use by ordinary people - There is no such thing as a manual for high-tech product or service that's written in plain language
			<p>Discomfort</p> <ul style="list-style-type: none"> - People are too dependent on technology to do things for them - Too much technology distracts people to a point that is harmful - Technology lowers the quality of relationships by reducing personal interaction
4.	Intention to Use (ITU)	Measures the intention of practitioners to use the Shariah document screening platform	<p>Assuming that the automatic Shariah document screening platform is available, I intend to use it.</p> <p>Given that I have access to the platform, I predict that I will use it.</p> <p>If the platform is offered in the market, I intend to use it.</p> <p>If the platform is available for subscription, I intend to propose to my company to use it.</p>

3.5.2 Sample Selection

This research adopts a stratified random sampling procedure to select participants among practitioners in Islamic financial industry. There are two steps in this procedure which first, the target population is divided into mutually exclusive and second, the sampling unit is chosen randomly from each subset (Churchill, 1999). This method is deemed suitable in the context of this research because of the availability of relevant data to characterize the target population.

Hence, this research followed the steps proposed by Stephan et al. (1940) for sampling selection. The population of industry practitioners was divided into two players which are Islamic banks and Takaful operators and the sample of practitioners were drawn from each of the players randomly. The use of stratified random sampling helps to ensure better sampling representatives. Moreover, according to Hair et al. (2003), this sampling method provides an opportunity to examine each sample group for an extensive comparison between the chosen groups. The sample of this research therefore coming from both Islamic banks and Takaful operators and they were either Shariah committee, Head of Shariah, or Shariah executives.

3.5.3 Sample size

There are factors that determine how large the sample size should be in research. These factors include model complexity, number of parameters, type of estimation method, and also analyses involved (Kline, 2011). Hence, for a multiple regression analysis, this research follows suggestion by Charles (2019) as in Figure 3.5, given 5 independent variables (k) and $\alpha = .05$ (significant level), a sample of 50 is sufficient to detect values of $R^2 \geq 0.23$. If the data is too small, it may be insufficient to represent the entire population (example, 20 samples). Therefore, the total number of industry practitioner involved in this research is 50. In addition, the population of Shariah advisors in Malaysia is less than 100 (bnm.gov.my).

Sample size requirements								
k	Significance level $\alpha = .01$				Significance level $\alpha = .05$			
	2	5	10	20	2	5	10	20
20	0.45	0.56	0.71	N/A	0.39	0.48	0.64	N/A
50	0.23	0.29	0.36	0.49	0.19	0.23	0.29	0.42
100	0.13	0.16	0.20	0.26	0.10	0.12	0.15	0.21
250	0.05	0.07	0.08	0.11	0.04	0.05	0.06	0.09
500	0.03	0.03	0.04	0.06	0.03	0.04	0.05	0.08
1000	0.01	0.02	0.02	0.03	0.01	0.01	0.02	0.02

Figure 3.5: Minimum sample for regression analysis (Charles, 2019)

3.5.4 Survey Analysis

Responses from the survey is analysed using a statistical software called SPSS. Few stages of analyses procedures were involved which includes data cleaning, reliability checking, crosstabulation, and multiple regression analyses. Results of the survey analysis is discussed in next chapter.

3.6 Chapter Summary

Overall, this chapter explains in detail the methodology adopted in this research. For document screening, Levenshtein Distance and FuzzyWuzzy Package is utilized to come out with a suitable algorithm. Meanwhile, survey is used to measure the intention to adopt the proposed Shariah Screening Platform (SHARSP) among financial industry practitioners.

CHAPTER FOUR

RESULTS AND DISCUSSION

4.0 Introduction

This chapter presents results of analysis of the proposed algorithms for document screening. First, proof of data pre-processing is shared and its followed similarity analyses algorithm by Leveinshtein and Fuzzy Wuzzy Package. This chapter further discusses results of the survey conducted among the industry practitioners which tackle their views on the proposed Shariah Screening Application (SHARSP). This chapter ends with a chapter summary.

4.1 Document Pre-Processing

This is an important and essential step prior to the next level of model development in NLP. A set of text corpus (data) collected from one or many sources, may have inconsistencies and ambiguity that requires pre-processing for cleaning it up. If text pre-processing is not done properly, it may affect the output of NLP model later. There were common text pre-processing procedures such as lowercasing, removing extra whitespaces, punctuations, stopwords, tokenization, spelling correction, stemming, and lemmatization. Using NLTK library, this research has combined few procedures of text pre-processing as below.

```
import nltk
import string
from nltk import word_tokenize
from nltk.corpus import stopwords
```

```

from nltk.stem.porter import PorterStemmer
from nltk import pos_tag

def preprocess(filename):
    f = open(filename,'r')
    text = f.read()
    text = text.lower()

    text_p = "".join([char for char in text if char not in
string.punctuation])

    words = word_tokenize(text_p)

    stop_words = stopwords.words('english')
    filtered_words = [word for word in words if word not in stop_words]

    porter = PorterStemmer()
    stemmed = [porter.stem(word) for word in filtered_words]

    pos = pos_tag(filtered_words)

    return words, filtered_words, stemmed, pos

```

4.2 Similarity Analyses

4.2.1 Levenshtein Distance

The first test of similarity is done using Levenshtein Distance. Given the algorithm written in Python, this research tests two different sets of text data where the first one is sample of Tawarruq product proposal (query) with Tawarruq policy document (corpus) and the second set is the same Tawarruq proposal (query) with Shariah Governance Policy Document (SGPD) as corpus. The second set of text is intentionally conducted to test the validity of Levenshtein Distance algorithm in checking similarity of two unrelated text/documents. Figure 4.1 and Figure 4.2 below present results of Levenshtein Distance for two sets of text.

checking for similarity using levenshtein

New Section

testing with related guideline-Tawarruq

Section

```
[ ] import Levenshtein as lev
String1 = "A tawarruq consists of two sale and purchase contracts. The first involves the sale of an asset by a seller (the Bank) to a purchaser"

options = {

"1.1": "Compliance with Shariah requirement is a prerequisite in ensuring the legitimacy of Islamic financial products and services. In meeting",
"1.2": "The Shariah contract-based regulatory policy is intended to ensure end-to-end compliance with Shariah and therefore, enhance the integrity",
"1.3": "The policy document contains two distinctive parts, namely the Shariah requirements and the operational requirements. The former highlights",
"1.4": "A tawarruq consists of two sale and purchase contracts. The first involves the sale of an asset by a seller to a purchaser on a deferred",
"1.5": "This policy document aims to:",
"1.5.a": "set out the Shariah rulings associated with tawarruq;",
"1.5.b": "set out key operational requirements with regards to the implementation of tawarruq;",
"1.5.c": "promote end-to-end compliance with Shariah requirements which include adherence to sound Islamic banking practices and safeguarding customer",
"1.6": "Part B sets out mandatory Shariah requirements to ensure the validity of the tawarruq as well as permissible optional practices.",
"1.7": "Part C provides operational requirements for a tawarruq on governance and oversight, structuring, risk management, financial disclosure,",
"1.7.a": "Principle 1: The IFI shall establish a comprehensive governance and oversight framework to ensure that a tawarruq is conducted based on",
"1.7.b": "Principle 2: The IFI shall ensure that the structuring and implementation of a tawarruq is supported by comprehensive policies and procedures",
"1.7.c": "Principle 3: The IFI shall implement a sound and integrated risk management system to effectively manage risks in line with the IFI's risk",
"1.7.d": "Principle 4: The IFI shall undertake a tawarruq in a fair and transparent manner in line with Shariah to protect stakeholder's interest;",
"1.7.e": "Principle 5: The IFI shall provide adequate disclosure and transparency to facilitate stakeholders' understanding and assessment of a tawarruq",
"2.1": "Subject to paragraph 2.2, this policy document is applicable to all IFIs as defined in paragraph 5.3.",
"2.2": "A licensed takaful operator is only required to comply with Part B of this policy document.",
"3.1": "The requirements in this policy document are specified pursuant to:",
"3.1.a": "sections 29, 57(1), 135(1) and 155 of the Islamic Financial Services Act 2013 (IFSA); and",
"3.1.b": "section 41(1) and constitutes a direction pursuant to section 129(3) of the Development Financial Institutions Act 2002 (DFIA).",
```

Table of contents

checking for similarity using levenshtein

New Section

testing with related guideline-Tawarruq

Section

+ Code + Text

Connect Editing

Distance	Ratio	Text
196	0.4380952380952381	"10.1" Each sale and purchase contract in a tawarruq is binding in nature and shall not be terminated or
190	0.4852233853006682	"10.2" The common inherent nature of each sale and purchase contract in a tawarruq is the transfer of ow
206	0.39572192513368987	"10.3" The sale and purchase contract in the tawarruq may take in the form of a murabahah or musawamah.
203	0.4	"11.1" The contracting parties to each sale and purchase contract in a tawarruq shall be a seller and a purchaser.
184	0.4766146993318485	"11.2" In relation to paragraph 11.1, the seller in the first sale and purchase contract shall not be th
196	0.45390970921985815	"11.3" The contracting parties in each sale and purchase contract in the tawarruq may enter into the sa
200	0.4326718816777042	"11.4" The contracting parties in the tawarruq shall be a natural person or a legal entity that has the
193	0.4390243902439024	"12.1" Each sale and purchase contract in a tawarruq shall be entered into through an offer and accepta
216	0.33766233766233766	"12.2" The offer and acceptance may be expressed orally, in writing or by any other methods recognised
221	0.3191489361702128	"12.3" The offer and acceptance referred to in paragraph 12.1 must be executed in the following sequence
197	0.4289044289044289	"13.1" An asset which is eligible to be used as the subject matter of a tawarruq shall be either a tangi
192	0.4547368421052631	"13.2" The asset which is the subject matter of a tawarruq may either take in the form of a whole asset
206	0.37688442211055273	"13.3" Pursuant to paragraph 13.2, the undivided asset shall be apportioned based on practices which ar
236	0.2642642642642643	"13.4" The following assets shall not be traded in a tawarruq:
201	0.4072398190045249	"13.5" The asset may be acquired by the seller through a purchase from a supplier, or through any other
198	0.41891891891891897	"13.6" In the event where the asset is not present at the majlis al-'aqd (contract session), the seller
194	0.44306418219461696	"13.7" Transfer of ownership from the seller to the purchaser shall take effect upon entering into a va
206	0.3914209115281502	"13.8" The purchaser shall take possession of the asset before the asset can be sold to a third party.
213	0.3685503685503686	"13.9" Possession of the asset shall either be in the form of qabd haqiqi (physical possession) or qabd
234	0.4771241830065359	"13.10" Possession of the asset shall take effect upon takhliyah (the seller releasing the asset to the
206	0.3877551020408163	"13.11" The seller shall be liable for any loss or damage of the asset before the purchaser takes posses
194	0.4893203883495145	"13.12" Any defect in the asset which occurred before entering into any sale and purchase contract in th
222	0.3202247191011236	"13.13" Where the situation in paragraph 13.12 occurs, the purchaser has the right to:
183	0.49438202247191011	"13.14" Any defect in the asset which is discovered and consented to by the purchaser at the time of ent
198	0.4056603773584906	"13.15" The contracting parties may mutually agree to specify the period of the defect option at the tim
190	0.43162393162393164	"13.16" Notwithstanding paragraph 13.12, the contracting parties may mutually agree for the seller to v
200	0.4260869565217391	"13.17" In the event that the contracting parties agree to the arrangement in paragraph 13.16, the waive
218	0.3370165745856354	"13.18" The same asset shall not be the subject matter of multiple tawarruq at any one time.

The least distance is: 1 The greatest ratio is: 0.9570093457943926

Tawarruq 1.4 A tawarruq consists of two sale and purchase contracts. The first involves the sale of an asset by a seller to a purchaser on a def

Figure 4.1: Results of Leveshtein Distance for related texts

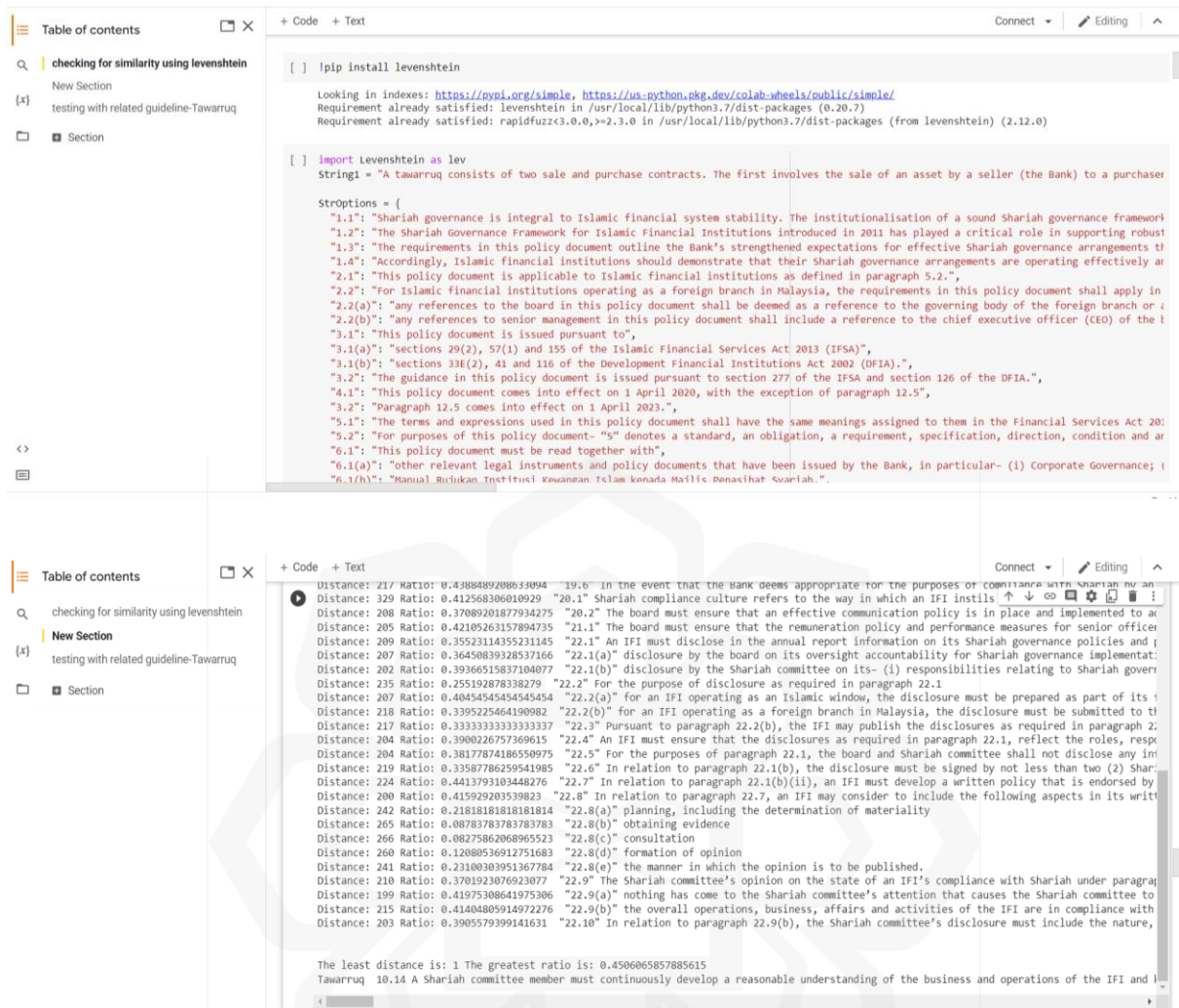


Figure 4.2: Results of Leveshtein Distance for unrelated texts

Based on the above similarity results, for the first set of related documents, the value of similarity is 0.957 which is near to 1.0 and can be interpreted as having minimum distance or high similarity level. Meanwhile, for the second set of unrelated texts, the value of similarity is 0.451 which is far from 1.0 and indicator of low level of similarity or unrelated text document.

4.2.2 FuzzyWuzzy Package

The second algorithm of measuring text similarity conducted in this research using Python programming language is FuzzyWuzzy Package. This technique is more on calculation ratio of distance between two texts (strings). Compared to Levenshtein Distance, FuzzyWuzzy may deal with more complex situations with few options of ratios. Figure 4.3 below presents results of FuzzyWuzzy Package.

```

Str1 = "A tawarruq consists of two sale and purchase contracts. The first involves the sale of an asset by a seller (the Bank) to a purchaser (C
Str2 = "A tawarruq consists of two sale and purchase contracts. The first involves the sale of an asset by a seller to a purchaser on a deferred
Ratio = fuzz.ratio(Str1.lower(),Str2.lower())
Partial_Ratio = fuzz.partial_ratio(Str1.lower(),Str2.lower())
Token_Sort_Ratio = fuzz.token_sort_ratio(Str1,Str2)
Token_Set_Ratio = fuzz.token_set_ratio(Str1,Str2)
print(Ratio/100)
print(Partial_Ratio/100)
print(Token_Sort_Ratio/100)
print(Token_Set_Ratio/100)
0.91
0.86
0.31
1.0

```

Figure 4.3: FuzzyWuzzy Package

Ratio of similarity is shown in Table 4.1 below. The value of exact matching ratio is high (0.91) represents the two set of texts (strings) 91 percent similar to each other. This is true when we read the sample texts in Figure 4.3 above.

Table 4.1: FuzzyWuzzy Ratios Value

No.	Ratio Similarity Checking	Purpose	Value
1	fuzz.ratio()	For exact matching, order of string does matter	0.91
2	fuzz.partial_ratio()	For partial matching, order of string does not matter	0.86
3	fuzz.token_sort_ratio()	Order of the words does not matter. It also ignores punctuation, and follows the concept of tokenization where the strings are converted into tokens and then sorted in alphabetical order for comparison analysis	0.31

4	fuzz.token_set_ratio()	Performs a set operation that takes out the common tokens (the intersection) and then makes fuzz.ratio(). Extra or same repeated words do not matter.	1.0
---	------------------------	---	-----

This method may also provide the string that have greatest ratio of similarity with the corpus tested as in Figure 4.5 below.

```

+ Code + Text
22.10": "In relation to paragraph 22.9(b), the Shariah committee's disclosure must include the nature, status and
}
max_ratio = 0
max_ratio_label = 0
max_ratio_label_content = ""
for nums in StrOptions.keys():
    Token_Set_Ratio = fuzz.token_set_ratio(String1,StrOptions[nums])

    print("Ratio:", Token_Set_Ratio/100, "", f"{nums}", StrOptions[nums])

    if max_ratio < Token_Set_Ratio/100:
        max_ratio = Token_Set_Ratio/100
        max_ratio_label = nums
        max_ratio_label_content = StrOptions[nums]

print('\n')
print("The greatest ratio is:", max_ratio, "\nTawarruq ", max_ratio_label, ":", max_ratio_label_content)

Ratio: 0.56 "13.3" An IFI must ensure that
Ratio: 0.21 "13.3(a)" majority of its Shariah committee members are Shariah qualified persons
Ratio: 0.28 "13.3(b)" the chairman of the Shariah committee must be a Shariah qualified person.
Ratio: 0.32 "13.4" Notwithstanding paragraph 13.2, an IFI operating as an Islamic window or a foreign branch in Malaysia must ensure that its
Ratio: 0.44 "13.5" Where there is more than one IFI within a financial group, section 30(2) of the IFSA states that one of the IFIs may submit
Ratio: 0.34 "14.1" An IFI shall appoint a secretariat to serve the Shariah committee. The secretariat has the responsibility to provide operat
Ratio: 0.27 "14.1(a)" coordinating communications and disseminating information among the Shariah committee, the board and senior management
Ratio: 0.24 "14.1(b)" performing in-depth research and studies on Shariah issues
Ratio: 0.35 "14.1(c)" providing day-to-day advice to relevant parties within the IFI on Shariah matters based on the rulings of the SAC and de
Ratio: 0.28 "14.1(d)" ensuring proper dissemination of decisions or advice of the Shariah committee within the IFI
Ratio: 0.34 "14.1(e)" undertaking administrative and secretarial functions to support the Shariah committee.
Ratio: 0.39 "14.2" An IFI must ensure that officers who perform the responsibilities in relation to paragraph 14.1(b) and (c) are officers wit
Ratio: 0.3 "14.3" An IFI must give the Shariah committee adequate time to deliberate all Shariah matters. In this regard, the secretariat has

+ Code + Text
Ratio: 0.21 "20.1" Shariah compliance culture refers to the way in which an IFI instils and promotes the behaviour
Ratio: 0.43 "20.2" The board must ensure that an effective communication policy is in place and implemented to advocate effective implementat
Ratio: 0.44 "21.1" The board must ensure that the remuneration policy and performance measures for senior officers responsible for ensuring S
Ratio: 0.37 "22.1" An IFI must disclose in the annual report information on its Shariah governance policies and practices. Such disclosures mu
Ratio: 0.33 "22.1(a)" disclosure by the board on its oversight accountability for Shariah governance implementation and the IFI's overall comp
Ratio: 0.45 "22.1(b)" disclosure by the Shariah committee on its- (i) responsibilities relating to Shariah governance; and (ii) opinion on the
Ratio: 0.25 "22.2" For the purpose of disclosure as required in paragraph 22.1
Ratio: 0.38 "22.2(a)" for an IFI operating as an Islamic window, the disclosure must be prepared as part of its financial institution annual r
Ratio: 0.33 "22.2(b)" for an IFI operating as a foreign branch in Malaysia, the disclosure must be submitted to the Bank.
Ratio: 0.31 "22.3" Pursuant to paragraph 22.2(b), the IFI may publish the disclosures as required in paragraph 22.1 on its website.
Ratio: 0.37 "22.4" An IFI must ensure that the disclosures as required in paragraph 22.1, reflect the roles, responsibilities and accountabil
Ratio: 0.31 "22.5" For the purposes of paragraph 22.1, the board and Shariah committee shall not disclose any information in the IFI's annual
Ratio: 0.26 "22.6" In relation to paragraph 22.1(b), the disclosure must be signed by not less than two (2) Shariah committee members.
Ratio: 0.34 "22.7" In relation to paragraph 22.1(b)(ii), an IFI must develop a written policy that is endorsed by the Shariah committee and ap
Ratio: 0.29 "22.8" In relation to paragraph 22.7, an IFI may consider to include the following aspects in its written policy to form an opinio
Ratio: 0.26 "22.8(a)" planning, including the determination of materiality
Ratio: 0.12 "22.8(b)" obtaining evidence
Ratio: 0.06 "22.8(c)" consultation
Ratio: 0.08 "22.8(d)" formation of opinion
Ratio: 0.27 "22.8(e)" the manner in which the opinion is to be published.
Ratio: 0.35 "22.9" The Shariah committee's opinion on the state of an IFI's compliance with Shariah under paragraph 22.1(b)(ii) shall be expre
Ratio: 0.37 "22.9(a)" nothing has come to the Shariah committee's attention that causes the Shariah committee to believe that the operations,
Ratio: 0.25 "22.9(b)" the overall operations, business, affairs and activities of the IFI are in compliance with Shariah but it has come to th
Ratio: 0.46 "22.10" In relation to paragraph 22.9(b), the Shariah committee's disclosure must include the nature, status and measures underta

The greatest ratio is: 0.94
Tawarruq 8.2 : A director must continuously develop and strengthen his knowledge and understanding on Islamic finance, as well as keep abreast

```

Figure 4.4: The Greatest Ratio of Similarity using FuzzyWuzzy Package

4.4 Results of Survey

4.4.1 Descriptive Analysis

This research managed to collect a total of 50 respondents from the industry. Descriptive analysis is performed to observe the background of the respondents. Out of 50, more than half of them were male (52%). Majority of the respondents were Shariah executive/officer (60%), followed by Shariah Committee (26%), and Head of Shariah Department (12%). In addition, one-third of the respondents currently serve full-fledged Islamic banks (36%) and the rest were from Takaful operators (24%), Subsidiary (22%), and Window Islamic banks (14%). Finally, in terms of job experiences, nearly half of them have served the industry between 1-6 years (48%) and the rest thirty six percent (36%) have more than 10 years experiences with the industry. Results of the descriptive analysis is presented in Table 4.1 until Table 4.4 respectively.

Table 4.2: Gender

		Frequency	Percent
Valid	Male	26	52.0
	Female	24	48.0
	Total	50	100.0

Table 4.3: Job Position

		Frequency	Percent
Valid	Shariah Committee	13	26.0
	Head of Shariah Department	6	12.0
	Shariah Executive/Officer	30	60.0
	Others	1	2.0

	Total	50	100.0
--	-------	----	-------

Table 4.4: Company

		Frequency	Percent
Valid	Full-fledged Islamic Bank	18	36.0
	Subsidiary Islamic Bank	11	22.0
	Window Islamic Bank	7	14.0
	Takaful Operator	12	24.0
	Others	2	4.0
	Total	50	100.0

Table 4.5: Work Experience

		Frequency	Percent
Valid	1-3 years	12	24.0
	4-6 years	12	24.0
	7-9 years	10	20.0
	more than 10 years	16	32.0
	Total	50	100.0

4.4.2 Multiple Regression Analysis

Multiple regression analysis is conducted to measure the effect of six independent variables towards Intention to Use the proposed Shariah Screening Platform called SHARSP. The independent variables are Perceived Usefulness (PU), Ease of Use (EOU), and Technology Readiness which presented by Optimism (OP), Innovativeness (IN), Insecurity (INS), and Discomfort (DIS). Table 4.5 below presents a model summary. The 'R' column represents multiple correlation coefficient where it measures the quality of the prediction of the dependent variables (PU, EOU, OP, IN, INS, and DIS). Thus, a value of 0.810 indicates a good level of prediction. In addition, 'R Square' which also called coefficient of determination explains the proportion of variance in the dependent variable (ITU) is explained by the independent

variables. As in Table 4.5, the value of R square is 0.657 indicates that all independent variables explain 65.7% of the variability of dependent variable which is intention to use the proposed SHARSP.

Table 4.6: Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.810 ^a	.657	.609	.21656
a. Predictors: (Constant), DIS, EOU, OP, IN, INS, PU				

Statistical Significant

Table 4.6 present the significant of the regression model tested. F-ratio in the ANOVA table becomes an indicator for overall model fit. The result shows that the independent variables statistically significantly predict the dependent variable (ITU) with $F(6, 43) = 13.713, p < .001$.

Table 4.7: ANOVA

ANOVA ^a						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3.858	6	.643	13.713	<.001 ^b
	Residual	2.017	43	.047		
	Total	5.875	49			
a. Dependent Variable: ITU						
b. Predictors: (Constant), DIS, EOU, OP, IN, INS, PU						

Accordingly, unstandardized coefficients indicate how much the dependent variable varies with an independent variables when all other independent variables are remained constant. Table 4.7 presents results of coefficients for each independent variables where if $p < 0.05$ it may conclude that the coefficients are statistically significance different from zero.

Overall, out of six variables, only three independent variables are statistically significant to influence the dependent variable (Intention to Use). They are Perceived Usefulness (PU), Innovation (IN), and Insecurity (INS). In details, the unstandardized coefficient for PU is equal to 0.366. These mean that for one level increase of perceived usefulness, there is an increase of 36.6 percent level of Intention to Use (ITU) the SHARSP among the practitioners. It is also same with the IN where the unstandardized coefficients is 0.218 indicates that one level increase of innovativeness, it will effect 21.8 percent of level of ITU. Hence, for the last significant variable, which is Insecurity, the result of coefficient is -0.686. This represents its negative effect towards intention to adopt the proposed platform by 68.6 percent.

Table 4.8: Results of Coefficients

Coefficients^a						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.781	.687		4.045	<.001
	PU	.366	.127	.434	2.872	.006
	EOU	.017	.066	.028	.257	.799
	OP	-.106	.126	-.101	-.838	.407
	IN	.218	.082	.350	2.666	.011
	INS	-.686	.202	-.368	-3.398	.001
	DIS	.081	.128	.069	.633	.530

a. Dependent Variable: ITU

A multiple regression was run to predict the Intention to Use (ITU) the proposed Shariah Screening Platform from Perceived Usefulness (PU), Ease of Use (EOU), and Technology Readiness which presented by Optimism (OP), Innovation (IN), Insecurity (INS), and Discomfort (DIS. Only three variables are statistically and significantly predicted ITU which are PU, IN, and INS with $F(6, 43) = 13.713$, $p < .001$, $R \text{ square} = 0.657$. All three variables added statistically significantly to the prediction, $p < .05$. It can be said that, the intention to use the SHARSP among the practitioners understudy is positively influenced by their perceived usefulness and sense of innovations. In addition, feeling insecurity will have a negative effect towards the intention to use SHARSP. In other words, the practitioners will use this platform when they feel it is secure for their job.

CHAPTER FIVE

CONCLUSION

5.0 Introduction

This chapter highlights the importance of the research findings with discussion with the market needs towards the proposed Shariah Screening Platform (SHARSP). Other than that, limitation and suggestion for future research are also shared in this chapter. This chapter ends with a chapter summary.

5.1 Discussion

This research is conducted to achieve two research objectives as below;

1. To develop the best algorithm for a digital assessment of Shariah document using Natural Language Processing (NLP)
2. To investigate views of industry practitioners on the proposed Shariah Screening Platform named as SHARSP

Adopting mixed-methodology approach, the first research objective is achieved via qualitative approach which related to data science. In details, a suitable algorithm is built for document similarity analysis using Python programming language. Two measures of similarity is adopted which are Levenshtein Distance and FuzzyWuzzy Package to compare two strings document. These algorithms then are used to develop a Shariah Screening Platform named as SHARSP. According to Kuruvilla (2021), among the advantages of these method are it may improve data quality, accuracy and is used for fraud detection within an organization. Among recent research

that have utilised this method is Bosker (2021) where his research has estimated the consistency, proficiency, and accuracy of fuzzy string matching (FuzzyWuzzy Package) as automated metrics of participants' accuracy in speech intelligibility. This method is also useful to record linkages (Wang et al., 2011), spelling inspectors, spam detector (Wei et al., 2009), for speech detection (Schalk & Zimmerman, 2005) and acoustic model training (Madan et al., 2020). However, it yet been explored in banking industry especially for document assessment. Levenshtein Distance is basically quantifying the match between a given string and a target string based on the number of shared characters. It is useful to quantify the accuracy of a particular document given the standard document or guideline by authority.

Accordingly, the second research objective adopts quantitative approach where data is gathered through a structured survey questionnaire. This survey is conducted online to save times and money. A total of 50 industry practitioners have involved in this survey which among them were Shariah committee, head of Shariah department, and Shariah executives. Based on multiple regression analysis, results show that among five factors that is tested to affect the technology adoption particularly SHARSP (the proposed platform), only two factors found to have positive significant effect towards intention to use SHARSP which are perceived usefulness, innovativeness. In addition, insecurity is found to have negative and significant effect towards technology adoption. There are a lot of studies that have adopted TAM and TRI in measuring acceptance towards technology among financial industry practitioners. Among them are Sellywati et al. (2022), Kamble et al. (2019), Kuo et al. (2013), and Martins et al. (2004). Overall, the practitioners understudy was in opinion that SHARSP would be useful for them and yet it has not been tested and this give an insignificant result of 'perceived ease of use' towards this platform. In addition, because of this platform is still in early stage and the

practitioners do not have a practical experience to use it, feeling of optimistic and discomfort are also found to have insignificant effect towards the usage of this platform.

5.2 Limitation

This research tries to propose an automated platform of document screening special for the Shariah department of Islamic banks and Takaful operators in Malaysia. It is important that all related document that involved in structuring new product is fully comply with the Shariah and the policy/guidelines by the Central Bank of Malaysia. There are few limitations of this research which first limited testing of algorithms. Due to time constraints, only two methods of document similarity are conducted in this research whereas there are a lot more algorithms offered in the market to cater the same objective. However, the two selected methods is deemed enough and suitable for the context of this research.

Second, limited number of respondents involved. This research only managed to get a total of 50 industry practitioners to answer the online survey forwarded to them. It is understood that although the survey is conducted online and require 10 minutes to finish, the practitioners have very tight schedule and limited time to answer for other matters.

Finally, real prototype of Shariah Screening Platform (SHARSP). The non-existence of real prototype is also among the limitation and factor that have influenced the results of multiple regression analysis (three independent variables are insignificantly affect the usage of the platform). It have also somehow affected the interest of the respondents to answer the online survey because of no real experience gained in that session. These limitations bring to the suggestion for future research which is discussed in next section.

5.3 Suggestion for Future Research

As highlighted in the previous section, this research would like to suggest the following for future research. First, in terms of algorithms building or selection. There are a lot of algorithms in the market that are readily available for adoption and testing in various context in the area of document matching. Future researchers may try to explore more algorithms either from the Python or other programming languages. This is important to ensure that the final algorithm is the best and suit with the context of study.

Second, the number of respondents should be increased. Future researchers may increase the number of respondents to utilise more advance statistical analysis and getting more accurate results. Method of data collection might be changed to interview approach to capture views or opinion of the regulator on the suggested automatic platform of document screening.

Finally, future research may extend this research by building a SHARSP prototype given an algorithm built in this research. It will give a good impression to the industry practitioners on a practical experience utilising the platform. This may later affect the results of survey and give a valuable input to the researchers to go for next level which is commercialisation of research output.

5.4 Chapter Summary

This chapter discusses output of this research, its limitations (data science, number of survey respondents, and SHARSP prototype), and suggestions for future researchers. It concludes this research.

REFERENCES

- Abu-Dalboah, H. M. (2013). A Questionnaire Approach Based on the Technology Acceptance Model for Mobile Tracking on Patient Progress Applications, *Journal of Computer Science*, Vol. 9 (6), pg 763-770.
- Ali, H., Abdullah, R., and Zaini, M. Z. (2019). Fintech and Its Potential Impact on Islamic Banking and Finance Industry: A Case Study of Brunei Darussalam and Malaysia, *International Journal of Islamic Economics and Finance*, Vol. 2 (1), pg 73-108.
- Al-Molijy, A., Hmeidi, I. and Alsmadi, I. (2012), Indexing of Arabic documents automatically based on lexical analysis, *International Journal on Natural Language Computing (IJNLC)*, Vol. 1 (1), pg. 1-8.
- Avillach, P., Joubert, M., and Fieshi, M. (2007). A Model for Indexing Medical Documents Combining Statistical and Symbolic Knowledge, *AMIA Annual Symposium Proceeding*, pag 31-35.
- Baeza-Yates R, Ribeiro-Neto B. (1999). *Modern Information Retrieval*. Addison-Wesley Longman; Harlow, UK.
- Benrazek, A. E., Kouala, Z., Farou, I., Ferrag, M. A., Seridi, H., and Kurulay, M. (2020). An efficient indexing for Internet of Things massive data based on cloud-fog computing, *Emerging Telecommunications Technologies*, Vol. 31(3), <https://doi.org/10.1002/ett.3868>
- Bigi, B. (2003). *Using Kullback-Leibler distance for text categorization*, Springer Berlin Heidelberg.
- Bosker, H. R. (2021). Using fuzzy string matching for automated assessment of listener transcripts in speech intelligibility studies, *Behavioral Research Methods*, Vol. 53 (5), pg. 1945-1953.

- Buttler, D. (2004). A short survey of document structure similarity algorithms, *International Conf. Internet Computing*, pg. 3-9.
- Chen, L., Li, X., Yang, Y., Kurniawati, H., Sheng, Q. Z., YunHu, H., and Huang, N. (2016). Personal health indexing based on medical examinations: A data mining approach, *Decision Support Systems*, Vol. 81, January, pg. 54-65
- Craswell N. (2009). Mean Reciprocal Rank. In: LIU L., ÖZSU M.T. (eds) *Encyclopedia of Database Systems*. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-39940-9_488
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology *MIS Q.*, Vol.13 (3), pg. 319-340.
- DuPree. (2015). Determinants of Intention to Use New Technology: An Investigation of Students in Higher Education, A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy In Information Systems College of Engineering and Computing Nova Southeastern University.
- Evans, C. (2015). Bitcoin in Islamic Banking and Finance. *Journal of Islamic Banking and Finance*, Vol. 3(1), pg. 1–11.
- Fox, C. (1989). A stop list for general text, *ACM SIGIR Forum*, Vol. 24, pg. 19-21.
- Gandhi, R. (2018). Support Vector Machine, Introduction to Machine Learning Algorithms SVM model from scratch, Retrieved at <https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47>
- Goker, A., and Davies, J. (2009). *Information Retrieval: Searching in the 21st Century*. John Wiley and Sons, Ltd., ISBN-13: 978-0470027622
- Gupta A., Verma, D. and Gambhir, K. (2012). Intelligent Information Retrieval, *Global Journal of Computer Science and Technology Software & Data Engineering*, Volume 12 (12).

Gupta D, Pratama M, Ma Z, Li J, and Prasad, M. (2019). Financial time series forecasting using twin support vector regression, *PLoS ONE* , Vol. 14 (3).

Han, J., Kamber, M., Pei, Jian. (2012). Getting to Know Your Data, Data Mining (Third Edition), The Morgan Kaufmann Series in Data Management Systems, Elsevier, pg. 39-82.

<https://blog.paperspace.com/implementing-levenshtein-distance-word-autocomplete-autocorrect/>

<https://data-flair.training/blogs/svm-kernel-functions/> Retrieved on 13th October 2020

<https://www.geeksforgeeks.org/major-kernel-functions-in-support-vector-machine-svm/>
Retrieved on 13th October 2020

Huang, A. (2008). Similarity measures for text document clustering. In Proceedings of the sixth New Zealand computer science research student conference (NZCSRSC2008), Christchurch, New Zealand (pg. 49- 56).

Indriyanto, I., and Sumitra, (2019). Measuring the Level of Plagiarism of Thesis using Vector Space Model and Cosine Similarity Methods, IOP Conf.

Jiffriya, M. A. C., Jahan, M. A., and Ragel, R. G. (2014). Plagiarism detection on electronic text based assignments using vector space model. In 7th International Conference on Information and Automation for Sustainability (pp. 1-5). IEEE.

Kadowaki, N., & Kishida, K. (2020). Empirical Comparison of Word Similarity Measures Based on Co-Occurrence, Context, and a Vector Space Model. *Journal of Information Science Theory and Practice*, Vol. 8(2), pg. 6-17.
<https://doi.org/10.1633/JISTAP.2020.8.2.1>

Kamble, S., Gunasekaran, A. & Arha, H. (2019). Understanding the Blockchain technology adoption in supply chains-Indian context, *International Journal of Production Research*, Vol. 57 (7), pg. 2009-2033, DOI: 10.1080/00207543.2018.1518610

- Ken Peffers, Tuure Tuunanen, Marcus A. Rothenberger & Samir Chatterjee, (2007). A Design Science Research Methodology for Information Systems Research, *Journal of Management Information Systems*, Vol. 24 (3), pg. 45-77, DOI: 10.2753/MIS0742-1222240302
- Kolade, O., Odumuyiwa, V., Abolfathi, S., Schröder, P., Wakunuma, K., Akanmu, I., Whitehead, T., Tijani, B., and Oyinlola, M. (2022). Technology acceptance and readiness of stakeholders for transitioning to a circular plastic economy in Africa, *Technological Forecasting and Social Change*, Vol.183.
- Kuo, KM., Liu, CF. & Ma, CC. (2013). An investigation of the effect of nurses' technology readiness on the acceptance of mobile electronic medical record systems. *BMC Med Inform Decision Making*, Vol 13 (88).
- Kuruville, (2021). A Comprehensive Guide to Fuzzy Matching/Fuzzy Logic, <https://nanonets.com/blog/fuzzy-matching-fuzzy-logic/>
- Laldin, M. A., & Furqani, H. (2019). FinTech and Islamic Finance. *Fintech in Islamic Finance: Theory and Practice*.
- Li, Jianping., Liu, Jingli., Xu, Weixuan., and Shi, Yong. (2004). Support Vector Machines Approach to Credit Assessment, • Conference: Computational Science - ICCS 2004, 4th International Conference, Kraków, Poland, June 6-9, 2004, Proceedings, Part IV, DOI: [10.1007/978-3-540-25944-2_115](https://doi.org/10.1007/978-3-540-25944-2_115)
- Lin, D. (1998). "An Information-Theoretic Definition of Similarity," *Proc. ICML*, pp. 296–304.
- Lovins, J. B. (1968). Development of a stemming algorithm, *Mech. Transl. Computer. Linguist.*, Vol. 11 (6), pg. 22–31.
- Madan, A., Khopkar, A., Nadig, S., Raghavan, K. M. S., Eledath, D., & Ramasubramanian, V. (2020). Semi-supervised learning for acoustic model retraining: Handling speech data

with noisy transcript. 2020 International Conference on Signal Processing and Communications (SPCOM), 1–5. 10.1109/SPCOM50965.2020.9179517

Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to information retrieval (Vol. 1, p. 496). Cambridge: Cambridge university press.

[Mark Scanlon, David Lillis](#) (2016). Conference: Proceedings of the 9th International Symposium on Digital Forensics and Information Security (DFIS 2016), Beijing, China

Martins, C., Oliveira, T., & Popovič, A. (2014). Understanding the Internet banking adoption: A unified theory of acceptance and use of technology and perceived risk application. *International Journal of Information Management*, Vol. 34 (1), pg. 1-13.

Maw, maw et al. (2020). Trends and Patterns of Text Classification Techniques: A Systematic Mapping Study. *Malaysian Journal of Computer Science*, [S.L.], V. 33, N. 2, P. 102-117, Doi: <https://doi.org/10.22452/Mjcs.Vol33no2.2>.

Mohamad Asmadi Abdullah, Rusni Hassan, Muhammad Naim Omar, Mohammad Deen Mohd Napiah, Ahmad Azam Othman, Mahamad Ariffin and Adnan Yusoff., (2014). A Study on the Shariah Decision Making Processes Adopted by the Shariah Committee in Malaysian Islamic Financial Institutions. *Aust. J. Basic & Appl. Sci.*, Vol. 8(13): pg. 670-675.

Parasuraman, A. (2000). Technology Readiness Index (TRI) a multiple-item scale to measure readiness to embrace new technologies. *Journal of service research*, Vol. 2(4), pg. 307-320.

Peppers, K., Tuunanen, T., Gengler, C. E., Rossi, M., Hui, W., Virtanen, V. & Bragge, J. (2006). The Design Science Research Process : A Model for Producing and Presenting Information Systems Research. In 1st International Conference, DESRIST 2006 Proceedings. (pg. 83-106). Claremont Graduate University.

Polettini, N. (2004). The vector space model in information retrieval-term weighting problem. *Entropy*, 1-9

- Saad, A., Alhabshi, S., Noor, A., & Hassan, R. (2020). Robo-Advisory for Islamic Financial Institutions: Shari'ah and Regulatory Issues. *European Journal of Islamic Finance*. <https://doi.org/10.13135/2421-2172/3992>
- Salton G. (1989). Automatic Text Processing: the transformation, analysis, and retrieval of information by computer. Addison-Wesley; Reading, MA.
- Schalk, T. B., & Zimmerman, R. S. (2005). Knowledge-based strategies applied to N-best lists in automatic speech recognition systems (United States Patent No. US6922669B2). <https://patents.google.com/patent/US6922669B2/en>
- Sellywati Mohd Faizal, Nahariah Jaffar & Azleen Shabrina Mohd nor (2022). Integrate the adoption and readiness of digital technologies amongst accounting professionals towards the fourth industrial revolution, *Cogent Business & Management*, Vol 9(1), DOI: 10.1080/23311975.2022.2122160
- Sharma, A. and Kumar, S. (2020). Bayesian rough set based information retrieval, *Journal of Statistics and Management Systems*, DOI: [10.1080/09720510.2020.1799575](https://doi.org/10.1080/09720510.2020.1799575)
- Shin, C. and Doermann, D. (2006). Document image retrieval based on layout structural similarity, *Int. Conf. Image Process. Comput. Vision, Pattern Recognit.*, pp. 606–612.
- Shouwei Li, Mingliang Wang, Jianmin He, "Prediction of Banking Systemic Risk Based on Support Vector Machine", *Mathematical Problems in Engineering*, vol. 2013, Article ID 136030, 5 pages, 2013. <https://doi.org/10.1155/2013/136030>
- Thakur, N., Mehrotra, D., Bansal, A., and Bala, M. (2019). Comparative Analysis of Ranking Functions for Retrieving Information from Medical Repository, *Malaysian Journal of Computer Science*, Vol. 32(1), p 18-30
- Thompson, V., Panchev, C. and Oakes, M. (2015). Performance Evaluation of Similarity Measures on Similar and Dissimilar Text Retrieval. In *Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, Vol. (1), pg. 577-584. ISBN: 978-989-758-158-8

- Van Rijsbergen CJ. (1979). Information Retrieval. Butterworths; London, UK.
- Wang, J., Li, G., & Fe, J. (2011). Fast-join: An efficient method for fuzzy token matching based string similarity join. Proceedings of the 27th International Conference on Data Engineering, 458–469. 10.1109/ICDE.2011.5767865
- Wei, C., Sprague, A., & Warner, G. (2009). Clustering malware-generated spam emails with a novel fuzzy string matching algorithm. Proceedings of the 2009 ACM Symposium on Applied Computing, 889–890. 10.1145/1529282.1529473
- [Wilkinson, R.](#) and [Hingston, P.](#) (1992), "Incorporating the vector space model in a neural network used for document retrieval", *Library Hi Tech*, Vol. 10(1/2), pp. 69-75. <https://doi-org.ezlib.iium.edu.my/10.1108/eb047846> Download as .RIS
- Witten IH, Moffat A, Bell TC. (1999). Managing Gigabytes. Morgan Kaufman; San Francisco, CA.
- Xie, S. Y., Liu, J. C., Liu and Han Wang. (2003). Research of intelligent information retrieval system based on three layers agent structure," Proceedings of the 2003 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.03EX693), Xi'an, 2003, pp. 2329-2332 Vol.4, doi: 10.1109/ICMLC.2003.1259897.
- Zubaidi, I. B., & Abdullah, A. (2017). Developing a Digital Currency from an Islamic Perspective: Case of Blockchain Technology. International Business Research, Vol. 10 (11), pg. 79.

APPENDIX 1 - LEVENSHTEIN DISTANCE

```
import Levenshtein as lev
String1 = "A tawarruq consists of two sale and purchase contracts. The first involves the sale of an asset by a seller (the Bank) to a purchaser (Customer) on a deferred basis. Subsequently, the purchaser of the first sale will sell the same asset to a third party on a cash and spot basis"

StrOptions = {
"1.1": "Shariah governance is integral to Islamic financial system stability. The institutionalisation of a sound Shariah governance framework strengthens public confidence in the integrity, management and business operations of the Islamic financial institutions.",
"1.2": "The Shariah Governance Framework for Islamic Financial Institutions introduced in 2011 has played a critical role in supporting robust and orderly developments of the Islamic finance industry in Malaysia. In particular, the framework has led to the institutionalisation of governance structures, policies and processes to promote end-to-end Shariah compliance in Islamic finance operations.",
"1.3": "The requirements in this policy document outline the Bank's strengthened expectations for effective Shariah governance arrangements that are well-integrated with business and risk strategies of the Islamic financial institutions. This augurs well with the growing scale and complexity of Islamic financial business, and recent policy developments in the areas of governance, compliance and risk management. Specifically, this policy document sets out strengthened oversight accountabilitys on the board, Shariah committee and other key organs involved in the implementation of Shariah governance. Another area relates to Shariah committee's objectivity to reinforce sound decision-making process and robustness of internal control functions for effective management of Shariah non-compliance risks.",
"1.4": "Accordingly, Islamic financial institutions should demonstrate that their Shariah governance arrangements are operating effectively and appropriate to their size, nature of business, complexity of activities and structure. The Bank also expects to see evidence of better alignment in promoting a strong Shariah compliance risk culture within Islamic financial institutions. This is in line with the more active roles expected of the board, Shariah committee and senior management.",
"2.1": "This policy document is applicable to Islamic financial institutions as defined in paragraph 5.2.",
"2.2": "For Islamic financial institutions operating as a foreign branch in Malaysia, the requirements in this policy document shall apply in respect of the Malaysian operations of the branch with the following modifications:",
```

```

    "2.2(a)": "any references to the board in this policy document shall
be deemed as a reference to the governing body of the foreign branch or
any of its committees locally",
    "2.2(b)": "any references to senior management in this policy documen
t shall include a reference to the chief executive officer (CEO) of the
branch and officers performing a senior management function in respect
of the branch operations.",
    "3.1": "This policy document is issued pursuant to",
    "3.1(a)": "sections 29(2), 57(1) and 155 of the Islamic Financial Ser
vices Act 2013 (IFSA)",
    "3.1(b)": "sections 33E(2), 41 and 116 of the Development Financial I
nstitutions Act 2002 (DFIA).",
}
min_distance = 1
max_ratio = 0
max_ratio_label = 0
max_ratio_label_content = ""
for nums in StrOptions.keys():
    Distance = lev.distance(String1.lower(),StrOptions[nums].lower())
    Ratio = lev.ratio(String1.lower(),StrOptions[nums].lower())
    print("Distance:", Distance, "Ratio:", Ratio, "", f"{nums}", StrOpt
ions[nums])

    if max_ratio < Ratio:
        max_ratio = Ratio
        max_ratio_label = nums
        max_ratio_label_content = StrOptions[nums]

print('\n')
print("The least distance is:", min_distance, "The greatest ratio is:",
max_ratio, "\nTawarruq ", max_ratio_label, max_ratio_label_content)

```

APPENDIX 2 – FUZZYWUZZY PACKAGE

```
Str1 = "As part of the Board's responsibility to promote sustainability growth and financial soundness of company, the Board must institutionalize a robust Shariah governance framework that is commensurate with the size, complexity and nature of Company's business."
```

```
Str2 = "As part of the board's responsibility to promote sustainable growth and financial soundness of an IFI, the board must institutionalise a robust Shariah governance framework that is commensurate with the size, complexity and nature of the IFI's business. The board's oversight accountability over Shariah governance and compliance must reflect the integration of Shariah governance considerations within the business and risk strategies of the IFI. In fulfilling this role, the board must-
```

```
Ratio = fuzz.ratio(Str1.lower(),Str2.lower())
Partial_Ratio = fuzz.partial_ratio(Str1.lower(),Str2.lower())
Token_Sort_Ratio = fuzz.token_sort_ratio(Str1,Str2)
Token_Set_Ratio = fuzz.token_set_ratio(Str1,Str2)
print(Ratio/100)
print(Partial_Ratio/100)
print(Token_Sort_Ratio/100)
print(Token_Set_Ratio/100)

.....
max_ratio = 0
max_ratio_label = 0
max_ratio_label_content = ""
for nums in StrOptions.keys():
    Token_Set_Ratio = fuzz.token_set_ratio(String1,StrOptions[nums])

    print(Token_Set_Ratio/100, f'"{nums}"', StrOptions[nums])

    if max_ratio < Token_Set_Ratio/100:
        max_ratio = Token_Set_Ratio/100
        max_ratio_label = nums
        max_ratio_label_content = StrOptions[nums]

print('\n')
print(max_ratio, f'"{max_ratio_label}"', max_ratio_label_content)
```

APPENDIX 3 SURVEY ONLINE FORM

Survey on Practitioners' View towards a Proposed Shariah Screening Platform

Assalamualaikum warahmatullah wabarakatuh.

Dear our respected respondent, this survey is conducted to get your valuable view on the proposed Shariah Screening Platform called SHARSP. This automatic platform is built based on Natural Language Processing (NLP) and Artificial Intelligent (AI) technology using Python programming language where the aim is to assist Shariah Department of Islamic banks and Takaful operators to run Shariah pre-screening process for the papers or product proposed.

A simple overview of the proposed platform is attached below. There are three main parts of the platform which are; 1) column for papers/proposal 2) Database which consists of Bank Negara policy document/guidelines/acts/International Standards 3) Result of Similarity/Shariah Compliancy.

We really hope to receive your feedback on the proposed Shariah Screening Platform especially in terms of perceived usefulness, ease of use, intention to use, and technology readiness. In case you have further suggestions for improvement, there is a column provided at the end of the survey.

For any enquiry, please do not hesitate to contact me at 011-62960996 (Dr. Marhanum CMS).

Barakallahu Fiikum for your valuable input.
Wassalam.

Research Team,
Dr. Marhanum Che Mohd Salleh
Dr. Rizal Mohd Nor
Kulliyah of Economics & Management Sciences
Kulliyah of Information & Communication Technology (KICT)
International Islamic University Malaysia

*Required

SHARIAH SCREENING PLATFORM (SHARSP)

Mahameri CHE MOHD SALIH | [User Info](#) | [Messages](#) | [Linktree](#) | [English](#) | [Community](#) | [Web](#) | [Logout](#)

SHARSP

Shariah Screening Platform

DOCUMENT SIMILARITY SEARCH ENGINE
Search or attach product proposal here

LIST OF SIMILAR DOCUMENT
Content of your proposal is found similar with this specific guidelines of Bank Negara Malaysia

SIMILARITY PERCENTAGE
The higher percentage show that the proposal is highly similar/relevant to the guidelines by Bank Negara Malaysia

Attachment

Click this button to show the similarity percentage

Print Report

PART A: BACKGROUND OF RESPONDENT

1. Gender

Mark only one oval.

- Male
- Female

2. Position in Islamic Bank/Takaful Operator

Mark only one oval.

- Shariah Committee
- Head of Shariah Department
- Shariah Executive/Officer
- Other: _____

3. Which player(s) that you serve currently.

Tick all that apply.

- Fullfledge Islamic Bank
- Subsidiary Islamic Bank
- Window Islamic Bank
- Takaful Operator
- Others

4. Experience in Islamic Financial Industry

Mark only one oval.

- Less than a year
- 1 - 3 years
- 4 - 6 years
- 7 - 9 years
- 10 years and above

5. Age Group

Mark only one oval.

- Below 30 years old
- 31 - 35 years old
- 36 - 40 years old
- 41 - 45 years old
- 46 - 50 years old
- 50 years old and above

6. Please rate your level of agreement on the statement below. *

Mark only one oval per row.

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
The Shariah document screening platform would assist the Shariah department to screen any Shariah non-compliance elements in the proposal.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The platform would shorten the time to prepare product proposal.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The platform would ensure the proposal is accurate according to the BNM guidelines/policy.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The platform would ensure the proposal having less mistake according to the BNM guidelines/policy.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Learning to utilise this platform would be ease for me.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would find it easy to use this platform.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The platform seems easy and understandable.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would find the platform to be flexible and user friendly	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

PART C: This part examines practitioners Intention to Use the proposed Shariah Screening Platform (SHARSP)

Please tick (/) on the appropriate box.

7. Please rate your level of confirmation on the statement below.

Mark only one oval per row.

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
Assuming that the automatic Shariah document screening platform is available, I intend to use it.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Given that I have access to the platform, I predict that I will use it.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
If the platform is offered in the market, I intend to use it.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
If the platform is available for subscription, I intend to propose to my company to use it.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

8. *Mark only one oval per row.*

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
Technology makes me more productive in doing my job	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
New technologies contribute to better quality of life	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Technology gives me more freedom of mobility	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
In general, I am among the first in my circle of friends to acquire new technology when it appears	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I can usually figure out new high-tech products and services without help from others	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I keep up with the latest technological developments in my area of interest	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Technical support lines are not helpful because they don't explain things in terms I understand	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sometimes, I think that technology systems are not designed for use by ordinary people	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
There is no such thing as a manual for high-tech product or service	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

that's written in plain language

People are too dependent on technology to do things for them

Too much technology distracts people to a point that is harmful

Technology lowers the quality of relationships by reducing personal interaction

9. In case you have any suggestion or comment to this survey or the proposed Shariah Screening Platform, please write it below. TQ

WE REALLY VALUE YOUR RESPONSE, COMMENTS, AND SUGGESTIONS. THANK YOU VERY MUCH.

This content is neither created nor endorsed by Google.