

# A Smart Application for Friday's Sermons Analysis

Abdulaziz A. AlShammari<sup>1</sup>, Tariq S. Almurayziq<sup>1</sup>, Yousef K. Qawqzeh<sup>2</sup>

<sup>1</sup>Department of Information and Computer Science, College of Computer Science and Engineering,  
University of Ha'il, Ha'il, KSA

<sup>2</sup>Department of Computer Science and Engineering, Hafr Al Batin University, Hafr AL Batin, KSA

## Summary

Regulations and rules are essential to all aspects of life and they suit both how one wishes to live, and how others should conform one's lifestyle. Several supporters have asserted that crime which is a violation against laws of the society, is integral to the human nature and hence the society can never be completely free from it. Modern society is characterized by increasing levels of risk discomposed by internal and external security threats. In this regard, security driven by technology is increasingly being used by government, corporate bodies, and individuals to monitor and reduce risks. This work proposed a system that will automatically suggest a sermon title for each Friday's sermon. The authorized Friday preachers (Imams) will receive the proposed title among its keywords in which they will be deployed into the sermon text. The system will then receive the uploaded sermon text to be analyzed for the purpose of accepting or rejecting the uploaded sermon by each and every Imam. The system will be utilized for awareness promotion based on security concerns in each area. It will utilize the analysis of Friday's sermon text by developing a customized machine-learning algorithm that discovers the hidden semantic structure of such documents using topic-modelling algorithm, in particular, the Latent Semantic Analysis algorithm. If the sermon scores more than 70% in deploying the keywords, it will be accepted, otherwise, it will be rejected.

**Keywords:** *Sermon Tracking; Sermon Management; Topic Modelling; Security Awareness.*

## 1. INTRODUCTION

Regulations and rules are essential to all aspects of life and it suit both how one wishes to live, and how others should conform one's lifestyle [1]. Several supporters have asserted that crime which is a violation against laws of the society, is integral to the human nature and hence the society can never be completely free from it. Modern society is characterized by increasing levels of risk discomposed by internal and external security threats. In this regard, security driven by technology is increasingly being used by government, corporate bodies, and individuals to monitor and reduce risk. Technology solutions could provide potential solutions to protect societies from hate, terrorist, and illegal sermons. Many methods have been used for raising the awareness of users, therefore, there is a need to customize some technological-

based methods to handle security issues raised by illegal and hate sermons. This work reviewed the related information and literature regarding security awareness and technological-based sermon orientation methods adoption. The main aim is to provide a proposed solution towards handling sermon delivery among Saudi society. In addition, the key challenges surrounding the successful implementation of the proposed security awareness system are addressed.

## Problem Statement:

As there are a variety of sermons that are received by the society inside Saudi Arabia, several security concerns are derived from the current situation. For example, some Imams are not committed to the proposed topics for Friday's sermon. Others may deliver hate, illegal, or unfocused messages to the audience. Security agencies inside Saudi Arabia need to monitor and control sermon message delivery using a reliable tool. In addition, up to my knowledge, there is no automated technological-based tool for handling such issues. Thereby, this research project tries to overcome such issues by introducing an ML algorithm that analyses sermon and produces guided instructions to Imams towards sermon delivery process.

## 2. LITERATURE REVIEWS

Today, the direction within the criminal justice system is clearly to improve the ability to share data among criminal justice agencies around the world [1-2]. According to [3], the state's digital agenda is closely linked to the digital transformation of businesses and governments, as well as the digital transformation of society. Information security (IS) and awareness raising (ISA) must be an integral part of these agendas. Many researchers emphasized the importance of thinking and awareness in general social sermons that affect safety issues and concerns. However, it is very important to make the sermon message interactive in order to effectively raise awareness. The importance of information security services and their

positive impact on society are being promoted and endorsed in many countries. However, [4] outlines the most important challenges to a successful implementation of information security awareness. Awareness can be raised by using ML that specializes in sermon analysis such as topic modeling. The field of topic modeling has become more important in recent years. Topic modeling is an unsupervised ML technique that allows you to organize text (or images, DNA information, etc.) to identify relevant text [5]. Machine learning (ML) is a field of artificial intelligence aimed at enabling machines to perform tasks skillfully using intelligent software [6-8]. Statistical learning methods form the backbone of intelligent software in which machine intelligence is developed. ML algorithms need data to learn, so the field must be related to the field of the database. There are also well-known terms such as knowledge discovery from data (KDD), data mining, and pattern recognition [1; 7]. The text mining studies are gaining more importance recently because of the availability of the increasing number of the electronic documents from a variety of sources. The resources of unstructured and semi structured information include the word-wide web, governmental electronic repositories, news articles, biological databases, chat rooms, digital libraries, online forums, electronic-mail, and blog repositories. Therefore, proper classification and knowledge discovery from these resources is an important area for research [9-10]. The documents representation is one of the preprocessing techniques that is used to reduce the complexity of the documents and make them easier to handle, the document has to be transformed from the full text version to a document vector. Text representation is the important aspect in documents classification, denotes the mapping of a documents into a compact form of its contents. A text document is typically represented as a vector of term weights (word features) from a set of terms (dictionary), where each term occurs at least once in a certain minimum number of documents. A major characteristic of the text classification problem is the extremely high dimensionality of text data. The number of potential features often exceeds the number of training documents. A definition of a document is that it is made of a joint membership of terms which have various patterns of occurrence. Text classification is an important component in many informational management tasks, however with the explosive growth of the web data, algorithms that can improve the classification efficiency while maintaining accuracy, are highly desired [11-13]. Scientists need powerful tools to read this data and understand the underlying themes of the data, such as humans. Topic modeling is a technique with a group of algorithms that reveal, discover, and comment on subject structures within a collection of documents [14]. Vector space models are the basis of many advanced information retrieval techniques [15] and subject models. VSM (Vector Space Model) was

the first simple algebraic model directly based on the term document matrix to extract semantic information from word usage [16]. The basic vector space model allows you to divide text into unigrams or bigrams of characters. It is based on the BOW approach, ignoring the exact order of terms in the document, but maintaining the frequency of occurrence of each term as an important factor [17]. VSM has extended its wings to a variety of applications, including ML, search engines, natural language processing, conceptual similarity, machine translation to measure context between words, phrases, and documents [18-19]. When modeling a search system such as a search engine or digital collection, not all words in a document are equally important. Therefore, assign weights to all terms in the document. This depends on how often the term appears in the document. The concept of weighting is to give more weight to unexpected events and less weight to common events. The topic modeling algorithm provides multiple perspective techniques for finding hidden semantics in document collections and grouping topics into topics. Latent Semantic Analysis (LSA) uses a term-by-term matrix with term weighting schemes such as Log Entropy and Term Frequency Inverse Document Frequency (TFIDF), and uses singular value decomposition to automatically index information from a range of objects. How to get it. Decomposition (SVD) method. LSA is one of the main problems of information retrieval technology, namely the processing of polysemous words, by assuming that the data is based on a potential semantic structure that is partially hidden by the randomness of word selection. LSA uses statistical techniques to estimate this latent structure and remove the masking "noise". In addition, LSA was regarded as a new general theory of similarity acquisition and knowledge representation that helped to simulate learning of vocabulary and other psycholinguistic phenomena [20].

### 3. METHODOLOGY

This work aims to monitor and control sermon delivery in Saudi Arabia by automatically provide a recommended topic for Friday's sermons. This work proposed a system that will automatically suggest a sermon title for each Friday's sermon. The authorized Friday preachers (Imams) will receive the proposed title among its keywords in which they will be deployed into the sermon text. The system will then receive the uploaded sermon text to be analyzed for the purpose of accepting or rejecting the uploaded sermon by each and every Imam. The system will be utilized for awareness promotion based on security concerns in each area. It will utilize the analysis of Friday's sermon text by developing a customized Machine-learning (ML) algorithm that discovers the hidden semantic structure of such documents using topic-modelling algorithm, in particular, the LSA algorithm. In addition, if the Imam missed the time

for uploading the sermon, or if he was unable to use the system due to any possible technical issues, the system then will provide the Imam with a ready sermon.

**Data collection**

This work requires the collection of qualitative data. Data sources for this project were determined and approved, and the needed coordination for data collection was made verbally as follows:

- Public Security
- Type of crimes

The Ministry of Islamic Affairs, Call and Guidance (MIADG):

- Mosque names
- Mosque locations
- Sermon database automatically

**Experiment Design**

In this phase, experiment design will assure the achievement of the previous requirements. As ML algorithms require inputs to be fixed length feature. This work will utilize one popular fixed-length feature for text analysis which is bag-of-words (Bow) method. However, this method completely ignores order and semantics of the words. According to the Bow approach, the words of interest such as "Dawaa", "Guide", "Salam", and "Pray".... etc., are equally distinct. Therefore, in the word embedding each word is expressed by a vector and similar words end up with similar vectors. Word embedding algorithms capture this similarity from the context of the word. Table 1 shows the main types of crimes according to security concerns.

Security Concerns	Legal Texts
<b>Informatics</b>	The blessing of the computer - the inviolability of defamation and defamation - not to infringe on privacy
<b>Administrative</b>	The prohibition of treason - the importance of honesty - the virtue of work - the prohibition of abuse of power - the obligation to keep a secret
<b>Violation of public security</b>	The blessing of security - the danger of corruption on earth - the

	importance of the unity of the ranks
<b>The mind</b>	The sanctity of alcohol - the danger of drugs - the blessing of contemplation - the blessing of the mind
<b>The offer</b>	The prohibition of adultery - the danger of being alone - the prohibition of mixing - the danger of immorality
<b>Self</b>	The sanctity of murder - the blessing of health - the provisions of assault
<b>The money</b>	The inviolability of usury - the blessing of money - the inviolability of theft - the inviolability of extravagance
<b>Weapons and explosives</b>	The danger of the weapon - the etiquette of arms - the harm of recklessness - the sanctity of assault
<b>Etiquette, worship and belief</b>	The virtue of literature - the importance of monotheism - the country of monotheism - the virtue of jurisprudence
<b>Terror</b>	The danger of terrorism - the prohibition of intimidation - the obligation to obey the guardian - the blessing of obedience and submission
<b>A thousand cases of loss</b>	The importance of the family - the importance of guidance - the provisions of the special mandate
<b>Transportation accidents</b>	The sanctity of recklessness - etiquette of the road - etiquette of the animal

Figure 1 illustrates the proposed algorithm for Friday's sermon classification development.

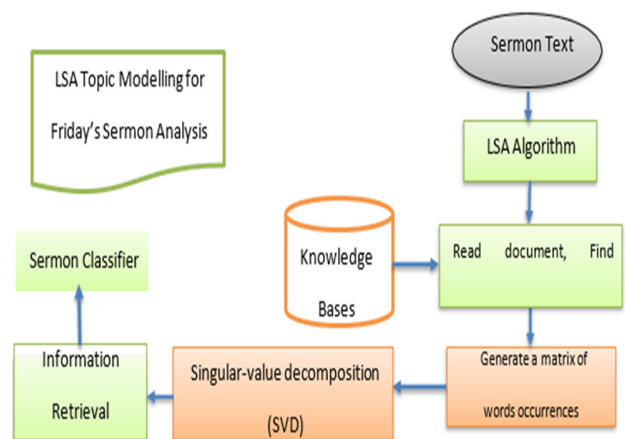


Figure 1. Friday's sermon algorithm flowchart

#### 4. RESULTS AND DISCUSSIONS

This work utilized a customized LSA algorithm in which the proposed application will suggest a topic for each and every Friday sermon. Each topic has certain key words in which it believed that these words if utilized in the text will deliver the needed message. For instant, an individual crime has several keywords related to this illegal action. In this regard, the LSA topic modeling will search the given text looking for the occurrences of the given keywords inside the text. The algorithm then counts these words and assigns an arbitrary weight for each word. Therefore, an overall weight for the used keywords in the text will be measured to measure the percentage of the overall utilization of keywords inside the text. Now, the proposed model can make a decision to accept or reject the text based on the achieved threshold. For this work, a threshold of 70% is recommended to judge the acceptance of any uploaded text to the system. If the uploaded text failed to achieve 70% utilization of the given keywords, it will be sent back to the Imam to re-inforce the keywords in the text. In some cases, were the Imam was unable to use the application and approve its own Friday sermon, the system will provide the Imam with an acceptable text. The application provides stats about the utilization of key words inside the texts and their impact on the overall intended message. The following demo demonstrates how the developed model analyzes the given text. It first shows the title of Friday's sermon along with its keywords that needed to be employed into the text. The authorized users, e.g. Imams will be able to access the application, read the title and its keywords, write the sermon, upload the sermon to the system, and then see the decision of the system in which it will be accepted or rejected. The following text is used as an example to test the developed model. In fact, we can use any text in Arabic language to run, test, and validate the developed system. However, this text is taken from a Friday's sermon:

أيها المسلمون "

حب المال والتعلق بطلبه والشغف بجمعه، والحرص على تمتيته، وداوم العمل على حراسته من الغوائل وكذا صيافته من الآفات مركز في الفطر، مستقر في العقول، مستحكم في النفوس

وإنما يكون صلاح هذا المال بحل أصله - وهو حديث بإسناد صحيح - وإنما يكون صلاح هذا المال بحل أصله، وطيب كسبه، ومشروعية مصدره، وهذا يستلزم التنزه عن أكل الحرام الخبيث الذي يبوء أكله بإثمه، ويكون وبالا عليه، جاء في الحديث - الذي أخرجه مسلم في "صحيحه"، والترمذي في "جامعه"، واللفظ لمسلم؛ عن أبي هريرة - رضي الله عنه - أنه قال: قال رسول الله - صلى الله عليه وسلم -: "إن الله طيب لا يقبل إلا طيباً، وإن الله أمر المؤمنين بما أمر به المرسلين، فقال: ﴿ يَا أَيُّهَا الرُّسُلُ كُلُوا مِنَ الطَّيِّبَاتِ وَاعْمَلُوا صَالِحًا إِنِّي بِمَا تَعْمَلُونَ عَلِيمٌ ﴾ المؤمنون: [51]، وقال: ﴿ يَا أَيُّهَا الَّذِينَ آمَنُوا كُلُوا مِن طَيِّبَاتِ مَا رَزَقْنَاكُمْ ﴾ البقرة: [172]. ثم ذكر الرجل يطيل السفر، أشعث أغبر، يمد يديه إلى السماء: يا رب، يا رب! ومطعمه حرام، ومشربه حرام، وغذي بالحرام، إلى ذلك يستجاب لذلك؟!"

Now after the authorized users logged to the system, they will be provided with the intended sermon title and its keywords. Figure ` below explains the control panel for the Imam. As shown in figure 1 below, the Imam is committed to write a sermon text that is suitable to the title and that employed the given keywords. Once the sermon text is ready, it will be uploaded to the system. Now the system will analyze the sermon looking for the employment of the keywords inside that text. As mentioned in the methodology section, we are looking for 70% employment percentage to accept the sermon. However, for testing purposes we reduce acceptance threshold to be 50% to test, verify and validate the develop approach. Now each and every keyword has its own weight. This weight reflects its importance inside the text. This weight varies depending on how much we think this given keyword is important in this place. For instance, keywords can hold variable weights to suit the topic, its importance, and its timing. The algorithm will search for each keyword, keywords will be determined in advance for each topic, and it will count the frequency of each keyword in the text. Based on that, a weight for each keyword will be calculated and recorded. Eventually, the total weight for all weights (each keyword has its own weight) will be used to judge the acceptance of the given sermon. We proposed the following equations to calculate the weight for each keyword and for the total weight. However, the weight equation is amendable.

$$\begin{aligned} Total\ weight_{keyword} &= weight\ of\ keyword \\ &+ (weight\ of\ keyword \\ &* weight\ of\ keyword) \\ &+ (weight\ of\ keyword \\ &* weight\ of\ keyword \\ &* weight\ of\ keyword) + \dots \end{aligned}$$

Now suppose that keyword 1 has a weight of 25%, and let us say it was found 3 times inside the text, then the weight of this keyword in the given text can be calculated as follow:

$$W_{keyword} = 0.25 + (0.25 * 0.25) + (0.25 * 0.25 * 0.25)$$

This equals to 0.25 + 0.0625 + 0.0156 = 0.328 which is equals to 32.8%

Now each keyword has a weight, the algorithm sums up all weights for all keywords and based on that if the overall weight exceeds the predetermined threshold, it will accept the text, otherwise it will be rejected.

#### 5. CONCLUSIONS

In this work, a text classifier model has been developed in which it is able to analyze sermons text, provide a decision to accept or reject the given sermon based on the predetermined threshold, and introduce a statistical report about the submitted sermon. The developed classifier utilized a customized LSA algorithm to implement text

classification in Arabic language. The main contribution of this work is the utilization of ML approach to classify Arabic text and provide a valuable service using mobile Android platform. The developed app allows the admin to give authorized accesses to Imams and other admins in which they can log into the system and perform certain functions based on their given permissions. The app will provide the Imams with the recommended sermon's title for each Friday prayer, along with the associated keywords of that title. The Imams will then write and prepare a sermon text in which it will be uploaded and checked by the developed classifier to be judge for acceptance or rejection. Another contribution of this work is that, certain promotional speeches can be assigned to certain areas or certain activities inside the kingdom. This facilitates the process of right and correct message delivery to the target audience. To conclude, this work utilized a customized LSA that reads, analyzes, and judge an Arabic text to be accepted or rejected for Friday's sermon. Moreover, if the sermon text achieved 70% of keywords deployment, it will be accepted.

## REFERENCES

- [1] M. Mohammed, M. B. Khan, and E. B. M. Bashie, *Machine learning: Algorithms and applications*, no. July, 2016.
- [2] R. Cioffi, M. Travaglioni, G. Piscitelli, A. Petrillo, and F. De Felice, "Artificial intelligence and machine learning applications in smart production: Progress, trends, and directions," *Sustain.*, vol. 12, no. 2, 2020, doi: 10.3390/su12020492.
- [3] O. Awodele, O. E. Ernest, O. A. Olufunmike, and S. O. Oluwawunmi Ugo-Ezeaba Anita A, "a Real-Time Crime Records Management System for National Security Agencies," *Eur. J. Comput. Sci. Inf. Technol.*, vol. 3, no. 2, pp. 1–12, 2015, [Online]. Available: [www.eajournals.org](http://www.eajournals.org).
- [4] M. Tamizharasan, R. S. Shahana, and P. Subathra, "Topic modeling-based approach for word prediction using automata," *J. Crit. Rev.*, vol. 7, no. 7, pp. 744–749, 2020, doi: 10.31838/jcr.07.07.135.
- [5] M. Scholl, "Social Media and Government Citizen Participation , Value Co-Creation and Service Delivery," vol. 9, no. January, p. 9981331, 2018.
- [6] M. Tamizharasan, R. S. Shahana, and P. Subathra, "Topic modeling-based approach for word prediction using automata," *J. Crit. Rev.*, vol. 7, no. 7, pp. 744–749, 2020, doi: 10.31838/jcr.07.07.135.
- [7] Qawqzeh Y, Alharbi MT, Jaradat A, Abdul Sattar KN. A review of swarm intelligence algorithms deployment for scheduling and optimization in cloud computing environments. *PeerJ Comput Sci.* 2021 Aug 25;7:e696. doi: 10.7717/peerj-cs.696. PMID: 34541313; PMCID: PMC8409329.
- [8] M. Alotaibi and W. Alfheid, "Information Security Awareness: A Review of Methods, Challenges and Solutions," *Internet Technol. Secur. Trans.*, no. March, pp. 119–127, 2018, doi: 10.2053/ICITST.WorldCIS.WCST.WCICSS.2018.0016.
- [9] B. Baharudin, L. H. Lee, and K. Khan, "A Review of Machine Learning Algorithms for Text-Documents Classification," *J. Adv. Inf. Technol.*, vol. 1, no. 1, 2010, doi: 10.4304/jait.1.1.4-20.
- [10] G. Kou, P. Yang, Y. Peng, F. Xiao, Y. Chen, and F. E. Alsaadi, "Evaluation of feature selection methods for text classification with small datasets using multiple criteria decision-making methods," *Appl. Soft Comput. J.*, vol. 86, p. 105836, Jan. 2020, doi: 10.1016/j.asoc.2019.105836.
- [11] W. Shang, H. Huang, H. Zhu, Y. Lin, Y. Qu, and Z. Wang, "A novel feature selection algorithm for text categorization," *Expert Syst. Appl.*, vol. 33, no. 1, pp. 1–5, Jul. 2007, doi: 10.1016/j.eswa.2006.04.001.
- [12] D. Blei, L. Carin, and D. Dunson, "Probabilistic topic models," *IEEE Signal Process. Mag.*, vol. 27, no. 6, pp. 55–65, 2010, doi: 10.1109/MSP.2010.938079.
- [13] M. R. Ghorab, D. Zhou, A. O'Connor, and V. Wade, "Personalised Information Retrieval: Survey and classification," *User Model. User-Adapted Interact.*, vol. 23, no. 4, pp. 381–443, Sep. 2013, doi: 10.1007/s11257-012-9124-1.
- [14] G. Salton, A. Wong, and C. S. Yang, "A Vector Space Model for Automatic Indexing," *Commun. ACM*, vol. 18, no. 11, pp. 613–620, Nov. 1975, doi: 10.1145/361219.361220.
- [15] P. D. Turney and P. Pantel, "From Frequency to Meaning: Vector Space Models of Semantics," *J. Artif. Intell. Res.*, vol. 37, pp. 141–188, Mar. 2010, doi: 10.1613/jair.2934.
- [16] P. Nakov and M. A. Hearst, "Solving relational similarity problems using the web as a corpus," *ACL-08 HLT - 46th Annu. Meet. Assoc. Comput. Linguist. Hum. Lang. Technol. Proc. Conf.*, no. June, pp. 452–460, 2008.
- [17] T. K. Landauer and S. T. Dumais, "A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge," *Psychol. Rev.*, vol. 104, no. 2, pp. 211–240, 1997, doi: 10.1037/0033-295X.104.2.211.
- [18] H. Chen, L. Xie, W. Feng, L. Zheng, and Y. Zhang, "Topic segmentation on spoken documents using self-validated acoustic cuts," *Soft Comput.*, vol. 19, no. 1, pp. 47–59, Jan. 2015, doi: 10.1007/s00500-014-1383-9.
- [19] C. Liu, H. Lin, S. Gong, Y. Ji, and Q. Liu, "Learning topic of dynamic scene using belief propagation and weighted visual words approach," *Soft Comput.*, vol. 19, no. 1, pp. 71–84, Jan. 2015, doi: 10.1007/s00500-014-1384-8.
- [20] P. Zhang, L. Wang, W. Huang, L. Xie, and G. Chen, "Multiple pedestrian tracking based on couple-states Markov chain with semantic topic learning for video surveillance," *Soft Comput.*, vol. 19, no. 1, pp. 85–97, Jan. 2015, doi: 10.1007/s00500-014-1375-9.