Received May 31, 2019, accepted June 5, 2019, date of publication June 17, 2019, date of current version July 11, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2923275

Halal Products on Twitter: Data Extraction and Sentiment Analysis Using Stack of Deep Learning Algorithms

ALI FEIZOLLAH[1], SULAIMAN AININ[1], NOR BADRUL ANUAR[2], NOR ANIZA BINTI ABDULLAH[2], AND MOHAMAD HAZIM[3]

[1]UM Halal Research Center, University of Malaya, Kuala Lumpur 50603, Malaysia
[2]Faculty of Computer Science and Information Technology, Department of Computer System and Technology, University of Malaya, Kuala Lumpur 50603, Malaysia
[3]Faculty of Computer Science and Information Technology, Department of Software Engineering, University of Malaya, Kuala Lumpur 50603, Malaysia

Corresponding author: Sulaiman Ainin (ainins@um.edu.my)

This work was supported in part by the Malaysian Higher Education Consortium of Halal Institutes, Ministry of Education, Malaysia, and in part by the University of Malaya under Grant MO001-2018.

ABSTRACT Twitter is a leading platform among social media networks. It allows microblogging of up to 140 characters for a single post. Owing to this characteristic, it is popular among users. People tweet about various topics from daily life events to major incidents. Given the influence of this social media platform, the analysis of Twitter contents has become a research area as it gives us useful insights on a topic. Hence, this paper will describe how Twitter data are extracted, and the sentiment of the tweets on a particular topic is calculated. This paper focusses on tweets of two halal products, i.e., halal tourism and halal cosmetics. Twitter data (over a 10-year span) were extracted using the Twitter search function, and an algorithm was used to filter the data. Then, an experiment was conducted to calculate and analyze the tweets' sentiment using deep learning algorithms. In addition, convolutional neural networks (CNN), long short-term memory (LSTM), and recurrent neural networks (RNN) were utilized to improve the accuracy and construct prediction models. Among the results, it was found that the Word2vec feature extraction method combined with a stack of the CNN and LSTM algorithms achieved the highest accuracy of 93.78%.

INDEX TERMS Twitter, algorithm, convolutional neural networks (CNN), long short-term memory (LSTM), recurrent neural networks, Halal tourism, Halal cosmetics, sentiment analysis.

I. INTRODUCTION

Sentiment analysis is a branch of natural language processing that analyzes text using machine learning algorithms. This method has attracted the attention of many developers and researchers. They identified polarity of text using sentiment analysis accurately. They have applied this method on various sources of text. Twitter is one of the sources that has been used to analyze sentiment.

The Twitter platform has been a key pillar of social networks. It is a podium for politicians, scientists, celebrities, etc. to express their views on a topic. As these sites are always accessible without the limitations of time and location, users regularly create contents ranging from daily life events to serious incidents. The influence of social media, in general, has become so large that even first-hand information on small and large incidents is gathered through social media platforms (Hu, Jamali, & Ester, 2012).

Twitter provides an application programming interface (API) to collect tweets. However, it is limited to the last seven days of data. A premium account, which costs hundreds of dollars, is required to access tweets that are older than seven days. In addition, Twitter has an advanced search feature that enables users to filter the desired data. Therefore, this work employs the advanced search function to collect tweets for the last ten years on halal products, i.e., halal tourism and halal cosmetics, in English and Malay languages. Given that many previous studies have focused on halal food, halal tourism and halal cosmetics are chosen as they are relatively new topics (Latiff, Masril, Vintisen, Baki, & Muhamad, 2019; Manan, Ariffin, & Maknu, 2019).

The associate editor coordinating the review of this manuscript and approving it for publication was Xi Peng.

The Halal industry covers a wide spectrum of topics. Research works such as Mostafa (2018), Wright and Annes (2013), among many works, have analyzed sentiment of users' opinion on Halal food. Besides Halal food, Halal tourism and Halal cosmetics are other topics in Halal industry that have gained popularity. In 2019, the New York Times reported that Halal tourism contribution to the global economy will jump to USD300 billion from USD180 billion Kamin [8]. Another report mentioned that Halal tourism is reshaping the global industry and many websites and applications have been emerging to accommodate needs for Halal tourism Belopilskaya [4]. Another interesting topic is Halal cosmetics, which is reported to reach USD54 million in 2022, compared to USD20 million in 2015 (Ray, 2017). The Halal cosmetics market has expanded its product base to prominently tap into the cosmetics market owing to increase in demand for halal cosmetic products worldwide.

Moreover, in recent years, many research works have been published on these two topics; however, there has been a lack of users' opinion analysis on these topics. For example, Battour *et al.* [3] explained the concept of Halal tourism and its future trends and challenges Battour and Ismail [2] explored the concept of Halal tourism along with the components that constitute the industry. They provided worldwide examples of some of the current best practices. The opportunities and challenges in developing and marketing Halal tourism were also discussed. In the topic of Halal cosmetics, research papers have tried to establish and explore its definition as well as future trends and challenges (Aoun and Tournois [1] and Swidi *et al.* [10]. The closest work to this paper was published by Majid *et al.* [10] in which they checked the relationship between awareness, religious belief and Halal product certification towards consumer purchase intention. They used a simple regression algorithm that is not as comprehensive as recent methods. Besides, their work did not analyze people's opinion.

There are many ways to process tweets. One of the most widely used methods is deep learning, which utilizes neural networks with several layers. This method has been used in many research areas. The reason for choosing deep learning is that it has produced satisfactory results and has not been used for halal tourism and halal cosmetics. It is then important to explore the performance of deep learning in such domains.

Therefore, in this work, we analyze tweets related to Halal tourism and Halal cosmetics. This work does not focus on the sentiment analysis algorithms instead our major contributions is to analyze a pool of 10-year tweets related to halal tourism and halal cosmetics, in English and Malay languages, and interpret their implications to the Halal industry.

The collection of tweets was performed by using Twitter's search function instead of API, which is costly and limited. Additionally, data collection and extraction in this work are part of our contributions as we used two languages to extract data. Moreover, our dataset is available upon request for future studies.

The remainder of this paper is organized as follows. In Section 2, we summarize related works. In Section 3, we explain the architecture and data collection and processing. Section 4 describes the experiments' setup and details. Section 5 presents the results of the experiments, and Section 6 concludes this work.

II. LITERATURE REVIEW

Social media networks allow users to create and share various types of contents and subsequently share them with other users. The contents can be created anytime and anywhere; therefore they offer freedom to users to create and share contents in real-time (Yoo, Song, & Jeong, 2018). Such characteristics have enabled researchers to analyze the content of social media to find patterns and trends, which can be used in many business areas. Accordingly, many studies (Bigné, Oltra, & Andreu, 2019 and Huertas & Marine-Roig, 2016) have been conducted to analyze the contents of social media.

A. COLLECTION AND EXTRACTION OF SOCIAL MEDIA POSTS

It is essential to correctly collect and extract data as they are used in the experiments to produce results. Many of the related works have used Twitter's API for data extraction, which is the conventional method. For instance, Bigné et al. (2019) analyzed tweets related to the tourism in Spain to study the hotel occupancy ratio and compared the results to official reports. They used Twitter's API for 17 days to retrieve related tweets. A similar work was published by Shayaa et al. (2018), where they analyzed the consumers' purchasing behavior and compared it with the consumer confidence index (CCI), which is published by the government. They collected Twitter data in English and Malay, only for two years. They used Twitter's API, which is costly and limited.

In another work, Yoo et al. (2018) analyzed tweets to detect events happening in an area, such as heavy snow and various incidents. They used 140 sentiment as their dataset and did not collect data themselves. Wang, Can, Kazemzadeh, Bar, & Narayanan (2012) analyzed tweets related to the US presidential election. They used Gnip Power Track, which is a Twitter data provider. This service still uses Twitter's API service to retrieve tweets. In another work, Si et al. (2013) retrieved 3-month data using Twitter's API and performed predictions on the stock market.

As opposed to the mentioned works, in this study we collected tweets based on the Twitter search function to avoid the cost of the API, which ranges from $149/month to $2,499/month, and its limitations on data requests, specifically number of requests. Then, we used an algorithm to filter the data based on language and extracted the posts in English and Malay for our analysis.

As sentiment analysis algorithms accept numbers as an input, there are mechanisms to convert text into a suitable format for them. These mechanisms are Word2seq and Word2vec. Xue, Fu, & Shaobin (2014) used sentiment analysis on the Sina Weibo to identify positive and negative contents. They used Word2vec technique to convert the data into input for algorithms. Details on these techniques are discussed in Section IV.

B. SENTIMENT ANALYSIS AND PREDICTION

Sentiment analysis is the process of evaluating a word or a sentence based on their sentiment. There are two approaches to perform sentiment analysis. One is based on a dictionary in which each word has a numerical value as polarity (Taboada, Brooke, Tofiloski, Voll, & Stede, 2011). The other approach is machine learning in which statistical methods are used to calculate the vectorized value of a word using word embedding. Then, the machine learning algorithm is trained using the digitized value of a word or a sentence. Some of the machine learning algorithms are support vector machine (SVM), random forest, naïve Bayes, convolutional neural network (CNN), recurrent neural network (RNN), and long short-term memory (LSTM) (Nasrabadi, 2007; Pedregosa et al., 2011). The machine learning approach is more popular nowadays due to its ability to learn and expand.

Yoo et al. (2018) collected tweets and utilized deep learning to study events in a location and to detect the possible location of incidents. They used CNN and LSTM algorithms for their experiments. In another work, Severyn and Moschitti (2015) analyzed sentiments of tweets using the CNN algorithm. They claim to have introduced a new method in sentiment analysis by combining CNN and an unsupervised algorithm. However, it is not accurate because the use of supervised algorithms leads to a more accurate results because of the labeling of data. However, this study uses two feature extraction methods along with a combination of CNN and LSTM to create a stack of deep learning algorithms to optimize the results.

III. ARCHITECTURE

This work aims at performing a sentiment analysis on two halal domains, namely halal tourism and halal cosmetics. The proposed architecture consists of data collection, data pre-processing, and data processing in the engine. Figure 1 shows the architecture in detail.

As mentioned earlier, the data collection of tweets is done using the Twitter search function. It uses the provided keywords to perform the search. The data pre-processing includes duplicate removals, retweet removals, and language detection.

The data then goes through vectorization, which prepares them for use in the data processing engine. This process is done by vectorizing words, which converts them into numbers, given that algorithms only work with numbers. Then, to utilize the data processing engine, it needs to be trained. This process involves gathering large datasets and training the algorithm on the collected datasets. This process outputs a model, based on which the tweets are grouped into positive or negative sentiment. The following sections discuss each module in detail.

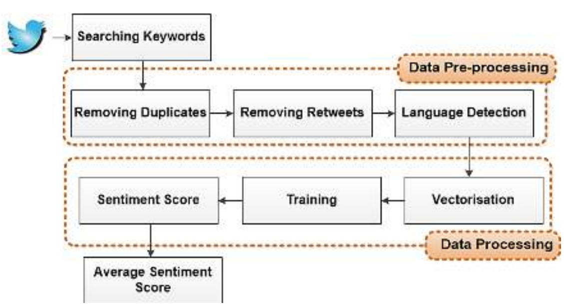

FIGURE 1. Architecture of the sentiment analysis process.

A. DATA COLLECTION

Our research team did data collection for this study. The data were collected from Twitter using keywords related to halal tourism and halal cosmetics from October 2008 until October 2018. Table 1 lists the keywords used for data collection. The keywords were determined after reviewing the literature on halal and after a brainstorming session among the researchers. A review of the literature illustrated that the term halal has been used interchangeably with Muslim friendly and Sharia thus, these keywords were added to extract the Twitter data.

The conventional method to collect tweets from Twitter is through their API, which allows developers and researchers to collect data. However, the API has many limitations, such

TABLE 1. Twitter keywords related to halal tourism and halal cosmetics.

Halal tourism	Halal travel	Halal accommodation
Muslim friendly tourism	Muslim friendly travel	Muslim friendly hotel
Muslim friendly trip	Sharia compliance accommodation	Sharia compliance hotel
Sharia compliance travel	Sharia compliance trip	Pelancong halal
Hotel halal	Lawatan halal	Penginapan halal
Halal cosmetics	Halal eyeliner	Muslim friendly cosmetics
Halal mascara	Sharia compliance cosmetics	Muslim friendly skincare
Halal makeup	Sharia compliance makeup	Muslim friendly lipstick
Halal lipstick	Sharia compliance lipstick	Muslim friendly makeup
Halal skincare	Sharia compliance skincare	Produk kecantikan halal
Muslim friendly accommodation	Muslim friendly accommodation	Muslim friendly accommodation
Kecantikan halal	Kecantikan halal	Kecantikan halal

TABLE 2. List of features returned from twitter.

Feature	Description
Full name	Name of the user
HTML	HTML code of the returned tweet
ID	ID of the tweet in Twitter database
Likes	Number of likes received for the tweet
Replies	Number of replies for the tweet
Retweets	Number of retweets for the tweet
Text	Main body of the tweet
Timestamp	Time and date of the tweet
URL	URL to the tweet
user	Username of the user who tweets

as limiting tweets up to seven days and a limited number of requests to the Twitter server. Therefore, we opted to collect data through the search feature of the Twitter website, using a Python script. This way, the limitations above are no longer an obstacle.

The search function returns the data listed in Table 2. Using this data, we were able to extract users' location by developing a Python script. This location is then used for analysis, which gives further insights into the collected data.

B. DATA PRE-PROCESSING

The data pre-processing process starts by removing retweets from our dataset. Retweets are just repetition of the original tweets, which confuses the algorithms in the sentiment analysis stage. Therefore, we remove retweets by identifying RT at the beginning of tweets.

Additionally, to identify duplicated tweets, the MD5 value of each tweet is calculated. MD5 is a hash function that returns a unique value for a given text. It is widely used in the computer security domain Rivest [15].

The search for keywords was done in two languages, English and Malay. As the Twitter search function returns posts regardless of their language, our collected data were multilingual. Therefore, we used Google's algorithm to detect the language of the text. This algorithm supports 55 languages, including English and Malay, and has 99% detection accuracy. The final dataset consists of 83,647 tweets in two languages that are not duplicated and are related to the halal tourism and halal cosmetic topics based on the keywords mentioned earlier.

C. DATA PROCESSING

After data collection and data pre-processing, the final dataset is fed into deep learning algorithms. This work takes advantage of stacked algorithms. In the next sections, we discuss deep learning and stacked algorithms in detail.

1) DEEP LEARNING

For several years, machine learning algorithms have helped in developing intelligent systems by training machines on how to make decisions. With a dataset labelled as input, machine learning constructs a model that is applied to new data to identify pattern similarities. Numerous studies have obtained significant results in the effective detection of intrusions by adopting a similar approach (Sangkatsanee *et al.* [16], Zhao *et al.* [19], Narudin et al., 2016, Feizollah *et al.* [5]). These algorithms have been around since the 1980s; however, the required computing power and data were not available until recently. With the increase in the amount of data and computing power in recent years, deep learning algorithms have re-surfaced.

Neural networks were originally proposed in the late 1940s but gained very little attention owing to the computationally expensive nature of the training necessary for all but the smallest networks. While there were some limited successes prior to the early 2000s, given the limitations of computational power, it was not practical to train neural networks with enough nodes and depth to produce useful results in most cases (Goodfellow *et al.* [6]).

This changed significantly around 2006 based on the general increase in computing power available and the use of GPUs in training. The realization that the training equations for neural networks could be run much more efficiently using the specialized hardware in GPU to handle the matrix multiplication operations resulted in ten times or more of increase in computation speed. This made it much more practical to train and employ neural networks for practical applications and led to the renewed interest currently seen.

Neural networks derive their name from the similarity of the individual computational components to neurons in the brain. Neurons or nodes in the network are connected by weighted edges, which allow for the flow of information through the network. At each neuron or node in the network, some form of non-linearity is applied to the summation of the weighted inputs from the incoming connected edges to produce an output. This output value is then propagated to other nodes in the network through the outgoing edges of the node.

The "deep" portion of the deep learning term comes from arranging the neurons in the network in layers of arbitrary size and then stacking these layers on top of each other. This approach effectively increases the computational complexity of the network. While it has been shown that a neural network with two layers is Turing complete, and thus capable of representing any function, an exponential cost would result if only the width and not the depth of the network is expanded (Goodfellow *et al.* [6]).

The major advancements in recent years improved the training of neural networks for practical applications. The typical method for training neural networks uses back propagation in combination with a technique such as stochastic gradient descent to update the weights of the edges and node biases. In back propagation, the output of the neural network is compared with the target or expected value, and then the difference between the two values is propagated backwards through all elements of the network. The error at each node is then used with an update algorithm, such as the stochastic

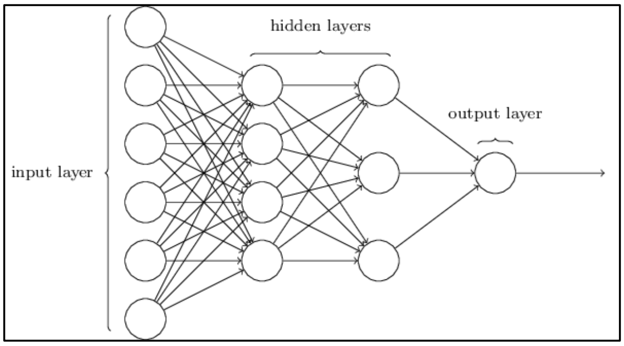

FIGURE 2. Structure of deep learning scheme.

gradient descent, to update the weights and biases in the neural network to minimize the training error. This process is repeated with additional training examples until the stopping criteria are met. There have been several modifications to the traditional stochastic gradient descent, such as Adagrad.

These modifications aim to improve the convergence by considering past variations in the gradient descent algorithm. This is done by taking the weight updates on individual parameters to smooth the convergence, usually by applying some form of dampening or momentum tracking. Additionally, in practice, the update procedure is typically conducted over a batch of training examples instead of individuals to speed up convergence. In batch processing, several training examples are provided at once, and the error is calculated across all examples when computing the updates. This can increase the processing performance in certain implementations and help to smooth the path of the learning algorithm by reducing the error introduced by reducing the effects of individual training examples (Goodfellow *et al.* [6]).

There are various types of deep learning available such as simple deep neural networks (DNN), RNN, LSTM, and CNN. CNN have recently gained popularity and success in the field of image recognition. Convolutional networks use a series of local filters, typically called receptive fields, applied across the image to generate a feature map.

The RNN and LSTM are proposed to deal with a series of related data. One of the appeals of RNNs is the idea that they can connect previous information to the present task, such as when using previous video frames might help to understand the present frame (Mikami, [11]).

The general and basic structure of the deep learning scheme is shown in Figure 2. It consists of an input layer, which is the input data to the algorithms; hidden layers, in which the algorithm makes numerous mathematical calculations; and the output layer, which is the result of the calculations of the algorithms.

As stated before, the concept of deep learning has been around for decades. However, it has gained momentum in recent years, for two reasons. First, the availability of massive data. Deep learning algorithms work well when dealing with a massive amount of data, such as credit card transactions or geospatial data. The more data is available, the better they are trained. In the case of this study, we gather multiple datasets for the training of the algorithms with 562,507 unique words.

The second reason is the availability of computing power. In recent years, the appearance of cloud computing has dramatically increased the computing power. It is possible to rent up to 64 cores of CPU from Google Cloud or Amazon AWS services at an affordable price. The deep learning algorithms can process terabytes of data using massive computing power to achieve high accuracy. Each neural network involves the following functions and calculations. A neural network follows the forward propagation in which the inputs are propagated across the layers. In addition, the network predicts the output based on inputs. Based on the explanations, the following functions are defined in the forward propagation.

$$\begin{aligned} z_1 &= xW_1 + b_1 \\ a_1 &= \tanh(z_1) \\ z_2 &= a_1W_2 + b_2 \\ y_2 &= softmax(z_2) \end{aligned}$$

The equations z_1 and z_2 are functions that take x as input and use W and b as weight and bias respectively. The ***tanh*** is an activation function that takes z_1 as input and passes the result to the next layer. In the output layer, the ***softmax*** function is used to calculate y_2, which is the prediction of the neural network.

2) STACKED DEEP LEARNING

This work uses stacking of the deep learning algorithms in the form of layers. Different algorithms are stacked up, and results of one algorithm are passed to another one. In this way, the weakness of one algorithm is compensated by the strength of the other algorithm. The experiments in the next sections show this advantage.

IV. EXPERIMENTAL DETAILS

This section explains several experiments that were conducted using the gathered dataset. It discusses the experimental details and architecture of the neural network.

A. FEATURE EXTRACTION TECHNIQUES

Two feature extraction techniques were used, as follows:

- Word2Seq: A word sequencing approach from the text without taking into consideration the weights of each word. This technique mapped the word sequence into a matrix with the length (input size) and height (number of observations).
- Word2Vec: A widely known feature extraction technique for text classification by using pre-trained Word2Vec (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013) or GLoVe models (Pennington, Socher, & Manning, 2014). The pre-trained models contain the weights of each word available inside the model. Thus, the main idea of Word2Vec is to supply the word sequence with weights of each word, making the embedding/input layer a vector representation of the texts.

These techniques were chosen as they are widely used and have resulted in acceptable outcomes (Xue et al., 2014).

B. DATASETS FOR MODEL TRAINING

To train the neural network algorithm, different datasets were used. These datasets cover a wide spectrum of topics. This is useful because the algorithm is trained comprehensively.

They are as follows:

- eRezeki (digital worker perceptions): This dataset is a private dataset collected from a crowdsourcing platform called eRezeki. It consists of 4,316 digital workers' perceptions and reviews about the platform. The reviews are in English and Malay languages. It includes a total of 3,475 unique words.
- IMDB (movie reviews): We collected this dataset from Maas et al. (2011). The dataset contains 50,000 movie reviews, split evenly between positive and negative reviews with 25,000 reviews each. It includes a total of 110,870 unique words.
- Amazon (product reviews): We obtained the Amazon reviews dataset from Feng & Zhu (2016) and McAuley, Targett, Shi, & Hengel (2015). It consists of 296,337 sports and outdoor product reviews. It includes a total of 134,758 unique words.
- Yelp (hotel and restaurant reviews): We obtained the Yelp reviews dataset from Zhang, Zhao, & LeCun (2015). The dataset is that used in the Yelp Dataset Challenge 2015. It contains 598,000 reviews from various hotels and restaurants across different countries. It includes a total of 313,404 unique words.

C. EXPERIMENTAL DETAILS

We experimented with stacks of algorithms and extraction techniques. It is important to present them so that the results of the experiments can be shown and compared with select the best results. Table 3 presents the details of each algorithm.

We experimented with different number of input sizes to select the most suitable input size that yields the best possible classification performance. Due to the difference in the underlying architecture of each model, the input sizes vary across different models. This is important as suitable number of input sizes allows the model to converge as efficient as possible during the training phase and reduces the degradation of model performance over training epochs. Therefore, each model is configured with different number of input size for optimal training and classification performance.

D. CONFIGURATION OF DEEP LEARNING ALGORITHMS

This experiment used the Keras library in Python to implement the neural network architectures. The reason for choosing Keras is the ease of implementation and modularity across different neural network architectures. The modularity reduces the complexity of building a powerful deep learning model. This allows us to focus more on feature extraction/generation and hyper parameter tuning rather than struggling to implement the neural network architectures.

The incredible flexibility of the Keras framework also allows us to easily combine different types of neural network layers to create a model that uses a combination of different neural network approaches in achieving the objective. Overall, three types of activation functions were used in this experiment.

- Hyperbolic tangent (Tanh): It works well with a simple recurrent layer by default. A simple-RNN layer is a fully connected RNN where the output is fed back into the input.
- Rectifier linear unit (ReLU): An LSTM layer works better with ReLU activation as compared with Tanh. ReLU can solve the vanishing gradient problem Ioffe and Szegedy [7] in back-propagated artificial neural network, which occurs in Tanh and Sigmoid activation functions.
    - Softmax function (Softmax): This is used in the final layer (fully connected) to convert logit scores (outputs) from the previous layer into probabilities that sum equals to 1. The output of the probabilities depends on the number of outputs/classes for the experiment. The number of outputs that are required to implement the Softmax activation function is more than one. Thus, Softmax is also used in multi-class classification settings.

Furthermore, the adaptive moment estimation (Adam) is used as the optimization algorithm for all the neural network models. Kingma and Ba [9] tested Adam on deep CNN and achieved a slightly higher performance as compared with SGD. Therefore, we chose to use it in these experiments. The other hyper parameters are:

- Learning rate = 0.01
- Decay rate = 0.000001
- Beta 1 = 0.9
- Beta 2 = 0.999 (set closest to 1 owing to the sparse-gradient problem)
- Epsilon = 1
- Decay = 0 (no decay rate implemented to provide faster convergence during training over a smaller number of epochs)

Different hyper parameters settings have been tested throughout the experiment to achieve optimum classification results.

E. LIBRARIES AND ENGINES

It is important to be able to replicate these experiments so that other researchers can benefit from them. Below, more details on our experiments and coding are presented.

- **Ubuntu 18.04** is used as the main OS because deep learning frameworks and architectures work better with Linux than with Windows (performance wise).
- **NVIDIA RTX 2080** (Gigabyte) is used as the main compute engine for the neural network.
    - 8 GB GDDR6 memory
    - 2944 CUDA cores
    - 368 Tensor cores

TABLE 3. Details of each algorithm.

Model	Training time (10 epochs)	Input Size	Layers
Word2Seq CNN	30 min	700	• 1 Embedding layer • 2 Convolution layers with max-pooling • 2 Fully-connected layers
Word2Seq LSTM	110 min	100	• 1 Embedding layer • 1 LSTM layer • 2 Fully-connected layers
Word2Seq CNN + LSTM	80 min	500	• 1 Embedding layer • 2 Convolution layers with max-pooling • 1 LSTM layer • 2 Fully-connected layers
Word2Seq CNN + Bi-RNN + Bi- LSTM	40 min	100	• 1 Embedding layer • 2 Convolution layers with max-pooling • 1 LSTM layer • 1 RNN layer • 2 Fully-connected layers
Word2Vec CNN	60 min	700	• 1 Embedding layer • 2 Convolution layers with max-pooling • 2 Fully-connected layers
Word2Vec LSTM	30 min	100	• 1 Embedding layer • 1 LSTM layer • 2 Fully-connected layers
Word2Vec CNN + LSTM	80 min	500	• 1 Embedding layer • 2 Convolution layers with max-pooling • 1 LSTM layer • 2 Fully-connected layers
Word2Vec CNN + Bi-RNN + Bi- LSTM	30 min	100	• 1 Embedding layer • 2 Convolution layers with max-pooling • 1 LSTM layer • 1 RNN layer • 2 Fully-connected layers

- **NVIDIA GPU Cloud (NGC)** Container for easy deployment of highly optimized Docker images for deep learning projects.
- **Docker v18.09.0** for hosting the NGC images.
- **NVIDIA Docker v2.0.3** for the customized docker manager for NGC images.
- **NVIDIA CUDA Toolkit v10.0** for the latest CUDA toolkit available to support next-gen Turing GPU.
- **Tensorflow:18.10-py3** Docker image from NGC (nvcr.io) for the latest version of highly optimized Tensorflow image using GPU learning.
- **Tensorflow 1.10** as the main backend of the neural network framework.
- **Keras v2.2.4** as the main high-level neural network API.

TABLE 4. Training results.

Algorithm	Accuracy	Precision	Recall	F-measure
word2seq_cnn	0.9323	0.9322	0.9323	0.9323
word2seq_cnn_birnn_bilstm	0.9242	0.9240	0.9242	0.9240
word2seq_cnn_lstm	0.9344	0.9344	0.9344	0.9344
word2seq_lstm	0.9223	0.9222	0.9223	0.9223
word2vec_cnn	0.9294	0.9292	0.9294	0.9292
word2vec_cnn_birnn_bilstm	0.9258	0.9256	0.9258	0.9256
word2vec_cnn_lstm	0.9378	0.9379	0.9378	0.9376
word2vec_lstm	0.9328	0.9328	0.9328	0.9326

V. RESULTS

Based on the prepared training datasets and the details mentioned above, Table 4 presents the training results of the algorithms.

Based on the gathered tweets, we quantified their sentiment using sentiment analysis algorithms. Basically, they first break a sentence into words and then associate each word with a pre-set sentiment value. Finally, they calculate the overall sentiment score of tweets. The sentiment score is either positive or negative. The actual score is in the range of 0 and 1. The algorithm calculated the scores of positivity and negativity. The final result depends on which score is higher.

The training results show that the stacks of CNN and LSTM algorithms achieved the highest accuracy, of 93.78%. This corroborates that stacks of deep learning algorithms result in a higher outcome. The training phase outputs a model

TABLE 5. Results of the experiments.

Algorithm	Average Sentiment Score		Average Sentiment Counts	
	Tourism	Cosmetics	Tourism	Cosmetics
word2seq_cnn	0.498424	0.509638	0.572451	0.58803
word2seq_cnn_birnn_bilstm	0.534376	0.556743	0.503484	0.643459
word2seq_cnn_lstm	0.51284	0.60334	0.599803	0.715858
word2seq_lstm	0.534981	0.544163	0.642145	0.637392
word2vec_cnn	0.422227	0.530906	0.502193	0.6232
word2vec_cnn_birnn_bilstm	0.313608	0.471898	0.38136	0.588955
word2vec_cnn_lstm	0.409511	0.552431	0.505102	0.662382
word2vec_lstm	0.543683	0.612374	0.640691	0.726964

that is used for sentiment analysis. Using this model and the collected tweets, the results are presented in the form of sentiment. The preliminary results of the experiments are presented in Table 5.

Table 5 includes the results of several experiments. The average sentiment score is calculated after gathering the sentiment score of all tweets based on Equation 1.

$$\text{Average Sentiment Score} = \frac{\text{sum of positive scores} - \text{sum of negative scores}}{\text{total number of tweets}} \tag{1}$$

Similarly, the average sentiment counts are calculated based on Equation 2.

$$\text{Average Sentiment Counts} = \frac{\text{number of positive tweets} - \text{number of negative tweets}}{\text{total number of tweets}} \tag{2}$$

The results show that the average sentiment score for tourism is ranged between 0.3136 by word2vec_cnn_birnn_bilstm and 0.5436 by the word2vec_lstm algorithm. This shows that the overall sentiment of the tweets toward halal tourism is positive. Similarly, word2vec_cnn_birnn_bilstm achieved an average score of 0.4718 and word2vec_lstm attained an average score of 0.6123. This means that the tweets for halal cosmetics show a positive sentiment. This also shows that the average sentiment score of halal cosmetics is higher than that of halal tourism. It indicates that people are more positive and more interested in the halal cosmetics issue. The results can be interpreted from the business aspect. Given that there is more interest in halal cosmetics, businesses can invest more in this domain. Moreover, sifting through the collected tweets shows that some users are interested in halal cosmetics in general, while others are interested in specific halal cosmetic products, such as halal nail polish.

TABLE 6. Results of the weighted sentiment.

Number of Weighted Positive Sentiments	Number of Weighted Negative Sentiments
90910	14632

As the results of all sets of metrics are available, the weighted average of each metric is calculated:

- Weighted average of positive sentiment probability:

$$\text{Weighted Average of Positive (WAP)} = \frac{\sum \text{positive probability}}{n}$$

- Weighted average of negative sentiment probability:

$$\text{Weighted Average of Negative (WAN)} = \frac{\sum \text{negative probability}}{n}$$

Therefore, after having weighted the sentiment probabilities, the sentiment polarity is identified using a set of rules:

$$\begin{aligned} \text{if} WAP &> WAN \text{ then sentiment} \\ &= \text{positive} \\ \text{else if } WAP &== WAN \text{ then sentiment} = \text{neutral} \\ \text{else sentiment} &= \text{negative} \end{aligned}$$

This reduces the possibilities of having tie conditions if only the predicted sentiments are considered. However, a tie condition is still possible, but the probability is low. Table 6 presents the weighted outcomes of the sentiment.

The weighted score is calculated on the whole dataset. It shows that the number of positive tweets is more than six times the number of negative tweets. This provides an overall view of users' sentiment toward halal tourism and halal cosmetics.

VI. DISCUSSION AND CONCLUSIONS

This work collected posts from Twitter on halal tourism and halal cosmetics for the last ten years in English and Malay languages. The data went through the pre-processing stage to remove retweets and duplicates. We also used a Python algorithm to detect the language of tweets and collected tweets in English and Malay. The vectorization was the next step to prepare the data for the algorithms. We used stacks of deep learning for sentiment analysis. An extensive collection of data was used to train the algorithms, which resulted in an accuracy of up to 93.78%. Then, the trained model was used to analyze the sentiment of tweets.

The results of the conducted experiments show that people's sentiment is positive toward halal tourism and halal cosmetics. The sentiment score of halal cosmetics was higher compared with that of halal tourism, which indicates more interest in this domain. This is a business opportunity that could yield profits in the future.

REFERENCES

[1] I. Aoun and L. Tournois, "Building holistic brands: An exploratory study of Halal cosmetics," *J. Islamic Marketing*, vol. 6, no. 1, pp. 109–132, 2015.
[2] M. Battour and M. N. Ismail, "Halal tourism: Concepts, practises, challenges and future," *Tourism Manage. Perspect.*, vol. 19, pp. 150–154, Jul. 2016.
[3] M. M. Battour, M. N. Ismail, and M. Battor, "Toward a halal tourism market," *Tourism Anal.*, vol. 15, no. 4, pp. 461–470, 2010.
[4] Y. Belopilskaya. (2019). *How Halal Tourism is Reshaping the Global Tourism Industry*. Accessed: May 15, 2019. [Online]. Available: https://hospitalityinsights.ehl.edu/halal-tourism-global-industry
[5] A. Feizollah, N. B. Anuar, R. Salleh, F. Amalina, R. U. R. Ma'arof, and S. Shamshirband, "A study of machine learning classifiers for anomaly-based mobile botnet detection," *Malaysian J. Comput. Sci.*, vol. 26, no. 4, pp. 251–265, Dec. 2013.
[6] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press. 2016.
[7] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," *Proc. 32nd Int. Conf. Int. Conf. Mach. Learn.*, vol. 37, Lille, France, Feb. 2015.
[8] D. Kamin. (2019). *The Rise of Halal Tourism*. Accessed: May 15, 2019. [Online]. Available: https://www.nytimes.com/2019/01/18/travel/the-rise-of-halal-tourism.html
[9] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014. Accessed Dec. 01, 2014, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980
[10] M. B. Majid, I. Sabir, and T. Ashraf, "Consumer purchase intention towards halal cosmetics & personal care products in pakistan," *Int. J. Islamic Marketing Branding*, vol. 1, no. 1, pp. 1–9, May 2015.
[11] A. MikamI, "Long short-term memory recurrent neural network architectures for generating music and japanese lyrics," Ph.D. dissertation, Boston College, Newton, MA, USA, 2016.
[12] M. M. Mostafa, "Mining and mapping halal food consumers: A geo-located Twitter opinion polarity analysis," *J. Food Products Marketing*, vol. 24, no. 7, pp. 858–879, Dec. 2018.
[13] F. Amalina, N. Feizollah, N. B. Anuar, and A. Gani, "Evaluation of machine learning classifiers for mobile malware detection," *Soft Comput.*, vol. 20, no. 1, pp. 343–357. Jan. 2016.
[14] A. Ray. (2017). *Halal Cosmetics Market by Product Type*. Accessed: May 15, 2019. [Online]. Available: https://www.alliedmarketresearch.com/halal-cosmetics-market
[15] R. Rivest, "The MD5 message-digest algorithm," Tech. Rep., 1992.
[16] P. Sangkatsanee, N. Wattanapongsakorn, and C. Charnsripinyo, "Practical real-time intrusion detection using machine learning approaches," *Comput. Commun.*, vol. 34, no. 18, pp. 2227–2235, Dec. 2011.
[17] A. Swidi, W. Cheng, M. G. Hassan, A. Al-Hosam, M. KASSIM, and A. Wahid, "The mainstream cosmetics industry in Malaysia and the emergence, growth, and prospects of halal cosmetics," College of Law, Government and Int. Studies, Universiti Utara Malaysia. Tech. Rep., 2010.
[18] W. Wright and A. Annes, "Halal on the menu?: Contested food politics and French identity in fast-food," *J. Rural Stud.*, vol. 32, pp. 388–399, Oct. 2013.
[19] M. Zhao, T. Zhang, F. Ge, and Z. Yuan, "RobotDroid: A lightweight malware detection framework on smartphones," *J. Netw.*, vol. 7, no. 4, pp. 715–722, 2012.

ALI FEIZOLLAH received the bachelor's degree in information system (IS) from the Ajman University of Science and Technology (AUST), Ajman, UAE, in 2010, and the master's degree in computer science and the Ph.D. degree from the University of Malaya, Kuala Lumpur, Malaysia, in 2011 and 2017, respectively, where he is currently a Postdoctoral Research Fellow. His research interests include mobile malware, intrusion detection systems, sentiment analysis, and machine learning.

SULAIMAN AININ received the M.B.A. and Ph.D. degrees from the University of Birmingham. She is currently with the University of Malaya Halal Research Centre, University of Malaya, Kuala Lumpur, Malaysia, where she is also the Dean of the Social Advancement and Happiness Research Cluster. Her research interests include organizational performance, information, computer and communication technology, social networks, and Halal tourism and cosmetics.

NOR BADRUL ANUAR received the master's degree in computer science from the University of Malaya, Kuala Lumpur, Malaysia, in 2003, and the Ph.D. degree in information security from the Centre for Security, Communications and Network Research, Plymouth University, U.K., in 2012. He is currently an Associate Professor with the Faculty of Computer Science and Information Technology, University of Malaya. He has published a number of conference and journal papers locally and internationally. His research interests include information security (intrusion detection systems), data sciences, artificial intelligence, and library information systems.

NOR ANIZA BINTI ABDULLAH received the bachelor's degree (Hons.) in computer science from the University of Malaya, Kuala Lumpur, Malaysia, the master's degree in interactive multimedia from Westminster University, London, U.K., and the Ph.D. degree in computer science from Southampton University, U.K. She has authored or coauthored over 50 refereed publications in international journals and book chapters. She has supervised several master's and Ph.D. degree's students with the University of Malaya. She also co-supervised several master's research students with the Moratuwa University of Sri Lanka. She is currently an Associate Professor with the Faculty of Computer Science and Information Technology, University of Malaya. Her research interests include personalized and adaptive learning, recommender systems, big data analytics, and content-based image/video retrieval. She serves as a Reviewer for several ISI-indexed journals and conferences.

MOHAMAD HAZIM received the M.C.S. degree (research) in opinion spam detection from the University of Malaya, Kuala Lumpur, Malaysia, where he is currently a Research Assistant. His research interests include opinion mining, social network analysis, and applied artificial intelligence.