Topic Identification for Filipino and English News using Bidirectional Long Short Term Memory with Attention Mechanisms

I. Introduction

News reports have served as the primary way for people to become aware of events significant to society, positive or negative. In the Philippines, there are different sources of news. These can be tabloids, newspapers, web articles, which can differ in language or dialect depending on their demographics.

Using *Natural Language Processing* (NLP), we can determine the topic being discussed within a news article. *Topic identification* is a study under NLP that deals with extracting context from an article and then determining the main focus of the said topic. State-of-the-art technologies such as *Recurrent Neural Networks* (RNN) are commonly used with NLP tasks, however, these technologies have the following flaws: RNNs are plagued with *Vanishing Gradient* problems, which at the worst case can prevent the RNN from learning through training. *Long Short-Term Memory* (LSTM), on the other hand, have overcome such limitations, but are limited in memory and cannot preserve dependency when there are more than a few steps.

However, machine interpretability in deep learning is quite challenging, deep neural networks can be considered as a black box model since it is *so deep* that it is difficult to interpret inside the network. Several machine interpretability techniques for deep learning are proposed by [9,10,11,12].

With such limitations in mind, another technology emerged: *Attention Mechanisms (AM)* [4]. The first use of this technology was for *Machine Translation* [4]. A few researches have studied the use of AM and have found out that this technology can be used for *Sentiment Analysis* and *Text Classification* [1, 3, 4].

The concept of Attention Mechanisms can be used on machine interpretability, coming from this, the objective of this study is to implement Attention Mechanisms for topic identification and compare its performance with LSTM implementations, Attention Mechanisms

are also used to interpret the prediction from the given input by visualizing which important words the model focuses to classify the given input. The data for the study will come from news articles from the web which are published by Filipino media companies. The language present in these articles are limited to Filipino and English only.

II. Statement of the Problem

The study of Topic Identification using Attention Mechanisms proposes a new system for machine interpretability of Long-Short Term Memory Networks. This is done by applying the Attention Mechanisms proposed by the research of [4] to Topic Identification. The broad problem of applying attention mechanisms to this problem branches out to the following more specific subproblems:

1. Will the new proposed system be accurately able to classify Filipino and English news using attention mechanisms?
2. What features/words are important to classify the given input?

III. Review of Related Literature

This project focuses on implementing a Bidirectional *Long Short-Term Memory* (LSTM) network with *Attention Mechanisms* (AM) to classify news articles. This section will feature NLP researches that worked with AM and some form of an *Artificial Neural Network* (ANN).

Starting with machine interpretability, deep neural networks are hard to interpret, there is a trade-off between accuracy and interpretability in deep learning models, *the deeper/complex the model is, the harder it is to interpret.* However, deeper learning models showed greater accuracy, but hard to interpret. The paper by [9,10,11,12] showed different techniques in interpreting deep learning models.

The approach of [1] to classify text utilizes a *Recurrent Neural Network* (RNN) model which was based on AM. The RNN takes in text as input, however, the text input can be quite long. The disadvantage of a vanilla RNN is that long text inputs cause it to create deeper layers that can cause the neural network to be inefficient. With an RNN with too many layers, the network will experience loss during information synthesis especially on the final stage of it. Too many layers can greatly hinder an RNN's ability to learn.

To overcome the said disadvantage when it comes to long text inputs, LSTM and *Gated Recurrent Unit* (GRU) variants of the RNN are implemented. The parameters of the nodes of these two can be adjusted so that the information that is being ran through the nodes can be controlled -- how much is passed on to other ones and how much is dropped can be fine tuned. LSTM contains more parameters while GRU contains lesser ones, which means faster processing.

The meaning of a word can be related to the words in front of it or behind it. Because of this, a Bidirectional RNN model must be implemented so this concept can be followed. In addition to this, AM is added to the model so that each word can have their own weight. Through

AM, the network learns to place heavier weights on keywords and lighter ones on non-keywords. This makes important features become more obvious. As mentioned before, LSTM and GRU implementations are going to be used. Both of these are Bidirectional.

The classifier is the second and final step. The *softmax classifier* is used for the classification of the text based on the feature vector produced by the previous step. A dropout algorithm is used to connect the said vector with the softmax classifier. In addition, the dropout algorithm prevents the model from overfitting, thus increasing its accuracy.

To test [1], two datasets are used for experimentation. First is a Chinese dataset from the 2014 Chinese Computer Society dataset. The second dataset is in English, publish by Reuters. These datasets are comprised of news articles and each dataset is split into two: train and test. After experimentation, the one using LSTM yielded 81.7% accuracy on the Chinese dataset and 46.4% on the English one. The GRU one produced an accuracy of 83.9% and 46% for the Chinese and English datasets, respectively. We can assume that this implementation is focused more towards the Chinese language.

[2] proposed a method called *Self-Attention Networks* (SANet) that models the interaction between word pairs. As the method's name suggests, the concept of *self-attention* is one of the main focus. Self-attention has been used in the field of machine translation and has produced state-of-the-art results.

SANet is derived from the *Transformer* architecture -- an architecture that utilizes self-attention and a neural network that doesn't make use of convolutional nor recurrent layers. According to the authors, the use of AM increases the interpretability of the rather hard-to-interpret neural networks. These neural networks are hard to interpret because of their hidden layers. SANet is different from Transformer because it utilizes a different form of attention and the output is not a sequence but a classification. Also, SANet employs a global max

pooling at the top of its architecture so that input of arbitrary length can be processed without truncating or padding. SANet accepts text input.

SANet first produces word embeddings for the text input, then processes the word embeddings using self-attention blocks, which is then passed to a feed-forward network comprised only of a single hidden layer with a rectifier (ReLU). The feed-forward network classifies the input using a softmax classifier.

[3] elaborates on the problem of Aspect-level Sentiment Analysis. If one is looking at a restaurant review that goes like *"the staff is nice, but the food is lacking"* and the focus is on the food, the review would be classified as a negative sentiment. However, if the aspect is on the staff, the review would be positive. The authors of this research claims that the use of neural networks in aspect-level sentiment analysis are still in infancy, and that attention is the effective mechanism to implement in this kind of problem.

The authors of [3] proposes an attention-based LSTM for aspect-level sentiment classification. The approach utilizes two concepts: *Aspect Embedding LSTM* (AE-LSTM) and *Attention-based LSTM* (AT-LSTM). AE-LSTM learns an embedding vector for each aspect while AT-LSTM utilizes AM to capture the key parts of a sentences, like we discussed in [1], however minus the bidirectionality. Each of these LSTM deals with one portion of the goal of aspect-level sentiment analysis. Finally, these two culminates as the *Attention-based LSTM with Aspect Embedding* (ATAE-LSTM) if their architectures are combined as a single network.

The dataset used by the experimentation of [3] is the SemEval 2014 Task4 dataset which is comprised of customer reviews. Each review has different aspects with their corresponding polarities. Two types of tests are ran: three-way classification, which is comprised of positive, negative, and neutral; and pos/neg classification which is comprised only of positive and negative. For the three-way classification, AE-LSTM and ATAE-LSTM yielded 68.9% and

68.7% accuracy. For the pos./neg. classification, AE-LSTM and ATAE-LSTM yielded 87.4 and 87.6. Both classification tests yield similar performance for both AE and ATAE.

[2] used [4] as a basis. The concept of *self-attention* and some of its advantages over convolutional and recurrent layers were discussed on the segment of [2]. [4] reasons out that the total computation complexity per layer is superior in self-attention compared to convolutional or recurrent layers. Another one is the length between long-range dependencies in the network. Longer distance between the paths can cause a harder time to learn dependencies. However, for the authors of this project, the interpretability offered by AM implementations is the best advantage over convolutional or recurrent approaches because words where the AMs are focused can be interpreted.

[5] tackles the problem of identifying whether an article is satirical or not. The research features a 4-level hierarchical model that makes use of encoders, attentions, and classifiers. To start of, the architecture starts with the character-level encoder to the word-level encoder. Both of these encoders utilize a CNN to encode word representation. It then uses paragraph-level attention to reveal which paragraphs contribute to decision making if a person to believe if what they are reading is satirical or not. The document-level classification classifies documents based on their linguistic features.

In addition to attention, [5] implements linguistic features to further determine accurately the satirical articles. As stated by the researchers, journalists and satirical writers display subtle habits when they write.

IV. Methodology

For the data, it is split into two datasets, the English news dataset, which came from [2.1] and the Filipino news dataset, which was manually gathered. The filipino news dataset is

gathered from news articles that are from the Manila Bulletin, Inquirer, ABS-CBN, and Philippine Star.

The system was written in the Python programming language. The researchers made use of its 3.6 release (Python 3.6.x). Several Python libraries are required to make the system work, such as Tensorflow, Keras, Pandas, and Scikitlearn. These libraries are the main modules used for the system. There are several other modules that are needed, but this section will only highlight the core ones.

A desktop computer was also used to train the model. It was equipped with a 6th-generation Intel i7-6700K processor, 16 gigabytes of RAM, and 2 Nvidia GeForce GTX 970 graphics cards. Google Colab [6] is also used as well since it provides a free Tesla K80 GPU.

Word Embeddings are the features used for this paper with a dimensionality of 300(a constant factor throughout). To reduce training time, pre-trained word embeddings are used. GloVE [7] Embeddings are used for the English news dataset and FastText [8] Word Embeddings are used for the Tagalog news dataset.

The total number for the English news dataset is 2,225 and the Filipino News Dataset is 1187. To determine the performance of the model, the basic accuracy metric is used. The model is fine-tuned by trying different numbers of LSTM units and different numbers of LSTM Layers.

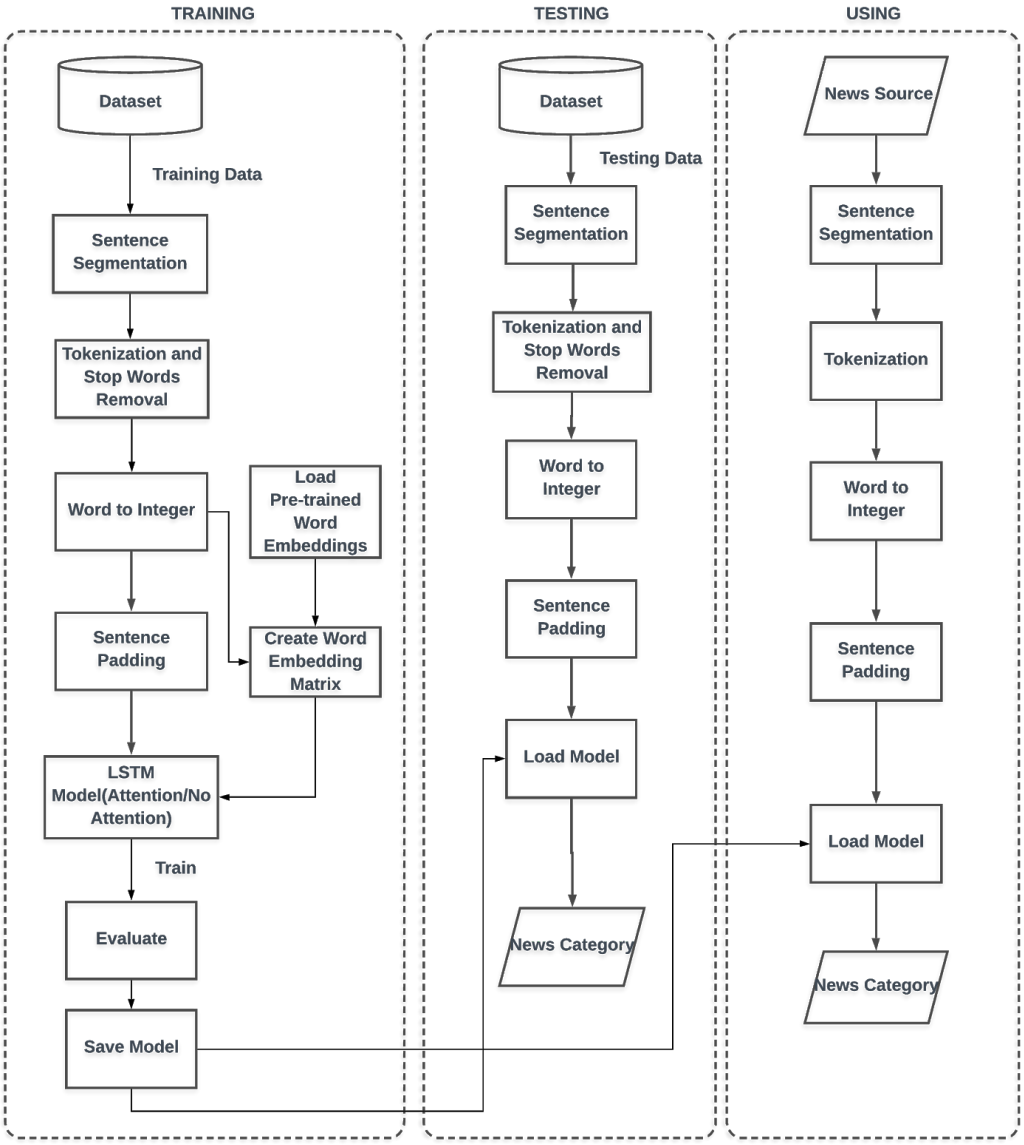


Figure 4.1. System Architecture

As seen on *Figure 4.1*, the system is comprised of three phases: *Training, Testing, and Using*. For the training part, this is where the AI model is trained that came from the dataset. A constant split of 80%-20% Training-Test split is used for both the English and Filipino news dataset.

For the testing part, this is where the system is evaluated basing from the accuracy, the higher the accuracy, the better performance of the system will be. Lastly, the Using part, this is

where the *deployment* of the model is used. The input could be any transcribed Filipino or English text from the newspaper or obtained through the internet.

For the input of the system, it can be any filipino or english news transcribed or obtained from the internet, the only limitations for this text is that the input should be in *Unicode* format since Python only reads files or texts through Unicode. For the training and testing phase, the input is obtained through the dataset. For the using phase, it can be outsourced from the internet or transcribed from a newspaper. The input is then tokenized, by splitting the paragraphs into sentences, then sentences into words. After tokenization, stop words are removed from the tokenized words to shorten the sentence length and remove unnecessary and unuseful words. This process is applied to training, testing, and using phase.

In the training phase, a word vocabulary or so-called *Dictionary* is created, this is used to create the word embeddings matrix. Each word is also mapped to an integer, where a one-to-one relationship is used. This is where the *word2index* dictionary is created and is used through the testing and training phase. Each word that is tokenized is then converted to an integer.

Finally, before the data is entered into the LSTM model, LSTMs only accept fixed-size sequences, in order to attain that, padding is used for all the sentences. To pad the sentences, the sentence who has the maximum length of N in all the training data is obtained. Then all the data, given their sequence length n_1, n_2, n_3, n_4 … n_i where *i* is the total number of data, their sequence length is extended to N. For example, if N=30, the maximum length of a sentence in the data, and a given training data *i* where n_i= 10, the sentence length of data *i* is extended to n_i=30.

After padding, then the data is then entered to the LSTM model, in this paper, 2 models are used, one with no attention mechanisms (*figure 2.2.a)*, and one model where attention mechanisms are present (*figure 2.2.b)*, which are shown in both the figures below, respectively.

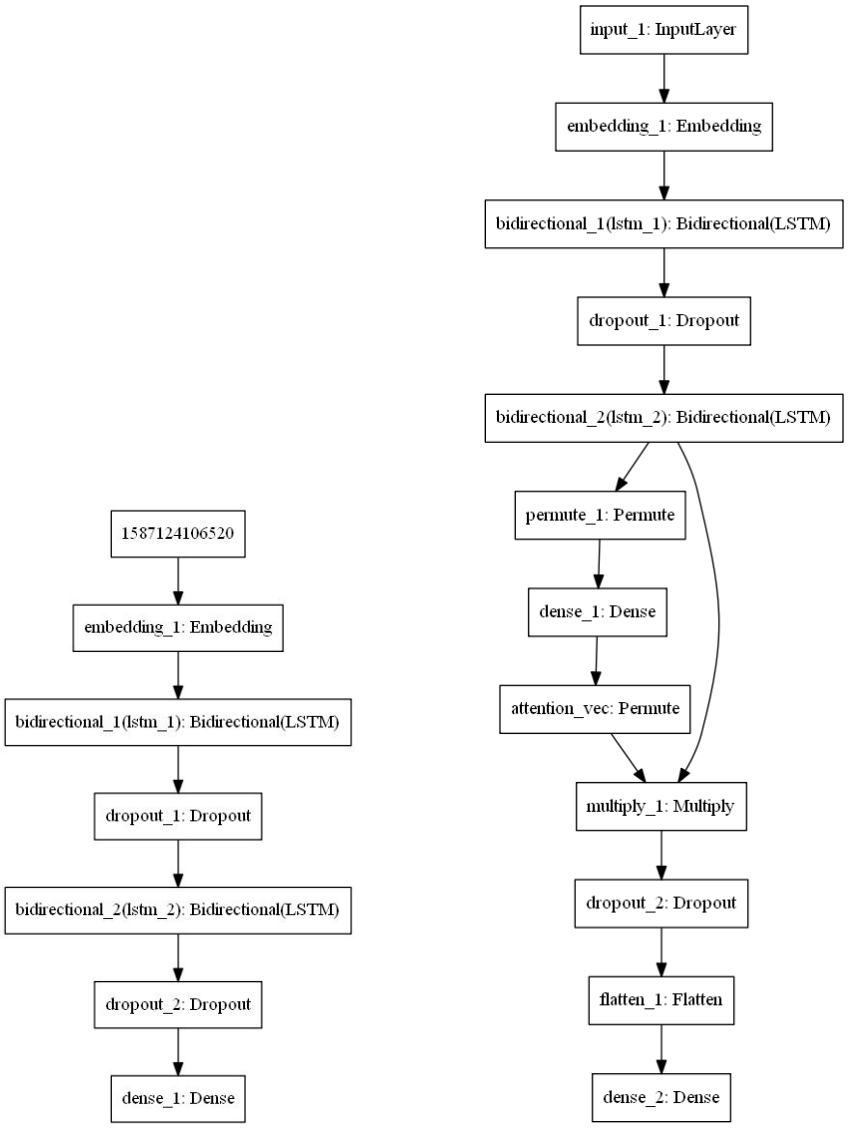


Figure 4.2.a (left). Architecture model for LSTM only

Figure 4.2.a (right). Model architecture for LSTM with Attention Mechanisms

The Attention Mechanisms is defined by [4] to be

$$Attention = \frac{Softmax(QK^T)V}{\sqrt{d_k}}$$

Where Q is the output of the last layer or the features extracted from the given input(if the Attention Mechanisms is the first layer) and K and V is the Key-Value pair derived from Q, to derive K and V, the Permute layer is used, then Q and K(which is transposed) are multiplied

using element wise dot product, then passed on to the Dense Layer for the Softmax function. The d_k refers to the dimensionality of the feature vector, which is 300 in this project. Lastly, the output from the Softmax Function are then multiplied to V using again the element wise dot product.

By using pre-trained word embeddings, trainable parameters are hypothesized to be decreased by a significant amount. The model is then trained by using the training data, and evaluated using the testing data. Once the accuracy is high enough, it is then saved so that it can be deployed on the Using phase.

V. Results and Discussion

Since there are two different models, there are also two tables, one table for LSTM with no Attention, and one table for LSTM with Attention Mechanisms. The accuracy metric is used since it is the most simple metric to define the performance of a given model. Moreover, both models is tested for 2 and 3 layers, with LSTM units tested for 100,128,256, and 300 units. Up to 4 LSTM layers are also planned but due to the large input of the dataset(2128 sequences for the English data set and 2012 sequences for the Filipino dataset), but due to time constraints, some model configurations are not tested.

After performing the experimentation, both attention and non-attention LSTM models yielded satisfactory results for the English and Filipino datasets. Table 5.1.a and 5.1.b shows the performance of both models when using the English dataset.

Table 5.1.a. Non-attention LSTM results for the English dataset

Number of LSTM Cells	Dropout	Number of LSTM Layers	Accuracy
100	0.3	2	0.9461883395
128	0.3	2	0.9551569509
256	**0.3**	**2**	**0.9707865**
300	0.3	2	0.9529147963

Table 5.2.b. Attention LSTM results for the English dataset

Number of LSTM Cells	Dropout	Number of LSTM Layers	Accuracy
128	0.5	2	0.9258426969
256	0.5	2	0.896629215
300	**0.5**	**2**	**0.96404496**

The highest accuracy for the non-attention network for the Filipino dataset is at 97.07%, on the other hand, the attention network produced a max accuracy of 96.40%, a difference of 0.67% between non-attention and attention models, respectively. The models on the Filipino dataset also produced a close gap in terms of accuracy. The non-attention model produced 83.61% while the one with AM produced 79.41%. The full details about the runs can be seen on Table 5.2.a and 5.2.b.

Table 5.2.a. Non-attention LSTM results for the Filipino dataset

Number of LSTM Cells	Dropout	Number of LSTM Layers	Accuracy
100	0.3	2	0.7773109239
128	**0.3**	**3**	**0.8361344543**
256	0.3	2	0.8235294123

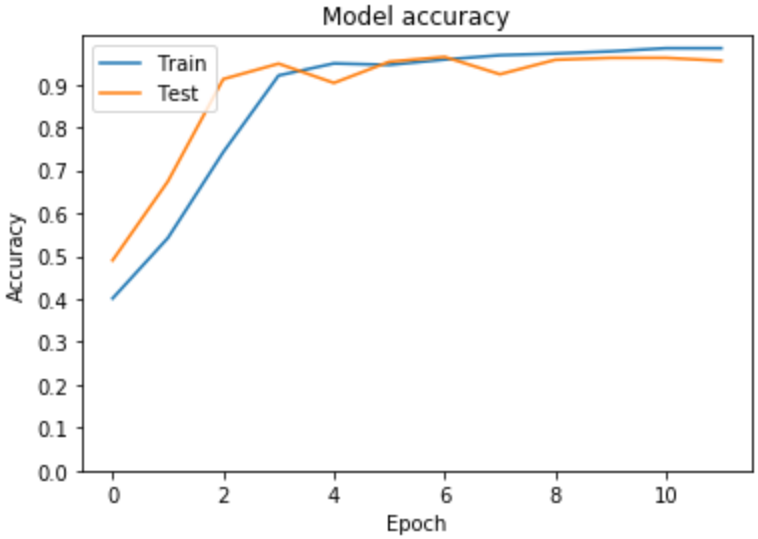
Table 5.2.a. Attention LSTM results for the Filipino dataset

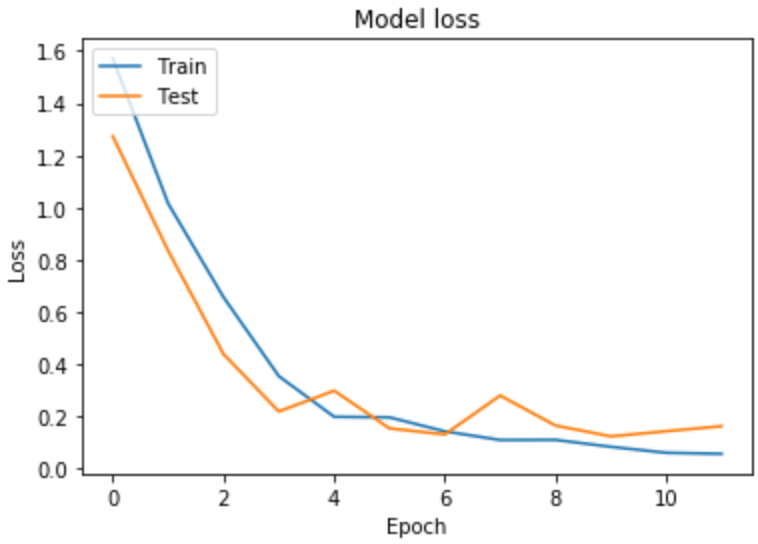
Number of LSTM Cells	Dropout	Number of LSTM Layers	Accuracy
128	**0.5**	**3**	**0.7941176**
256	0.5	3	0.64285713
300	0.5	3	71.8487365

More LSTM layers are added in the Filipino dataset, this is due to the fact that the data is more complex, and not standard, and the results provided a much lower accuracy compared to the English dataset. And the difference in the Filipino News between Attention and Non-Attention models are around 4%. Compared to the English dataset, which is only 0.67%, however, adding attention mechanisms may add model interpretability to know how the model behaves in a given input.

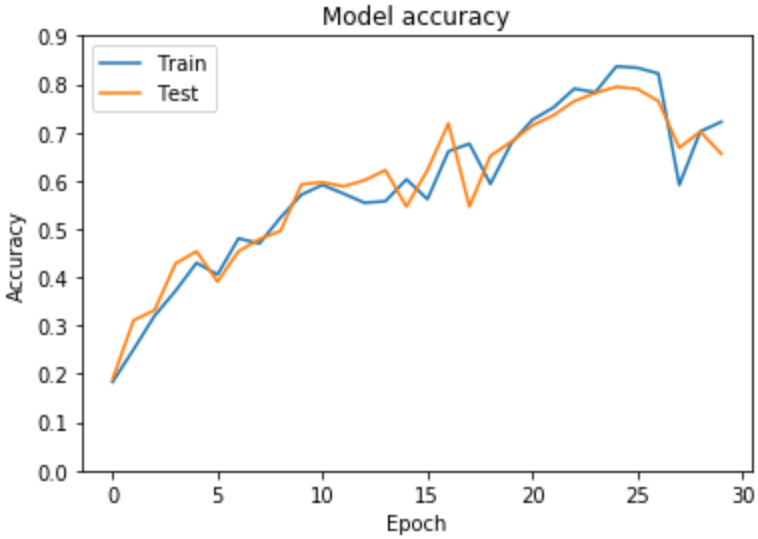
Adding visualizations to the comparison, a graph between Attention Mechanisms of the English Dataset and Filipino Dataset is shown below:

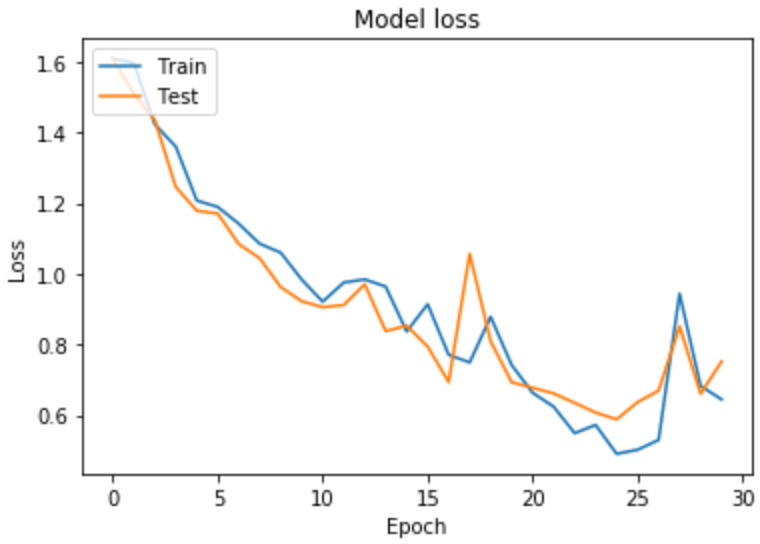
English Dataset:





Filipino Dataset:





The model loss visualizes the *gradient descent* of the models, and the accuracy visualizes the accuracy rising between epochs. For the english dataset, it is seen that the descent of the loss function and the rise of the accuracy is very smooth for 2 layers. For the filipino dataset, the loss function and the accuracy is fluctuating between their descent and ascent, respectively despite having 3 layers of LSTM. One reason for that is that the data may be too complex for the model. One recommendation would be trying 4 layers and 5 layers, however due to time constraints, 4 and 5 layers is not experimented in this project.

A sample input is shown below to demonstrate how Attention Mechanisms interpret a given input outside the dataset and obtained from the internet:

Input Text:

Disney/Marvel's Avengers: Endgame has moved up the offshore chart again, overtaking Avengers: Infinity War as the No. 3 highest grossing movie ever at the international box office. Friday's overseas take was $88.2M, lifting the running cume to $1.4B. With the domestic Friday included, Endgame has a global total of $1,914.5M as it hammers its way to crossing the $2B mark by the end of today, a milestone we noted yesterday would be the case. Already through Thursday, Endgame became the No. 5

highest grossing global movie of all time, having passed Jurassic World. The next films the assembled Avengers have in their sights on the worldwide chart are Infinity War and Star Wars: The Force Awakens which will likely be toppled through Sunday. Internationally, Endgame is looking to soon dunk Titanic and become the No. 2 film ever overseas Not included in the above Friday figures is the Saturday China estimate which is coming in around $21M. That's a 41% drop from Friday and a 74% drop from last Saturday which was heavily frontloaded. In both cases, the Saturdays fall ahead of a workday Sunday which is intended to make up for the extended holiday period the market had this week. The local cume with Saturday's estimate is $568.4M. Ticketing platform Maoyan continues to jockey its ultimate projection, today landing on RMB 4.08 ($606M). This Anthony and Joe Russo-helmed extravaganza has had a fantastic run through its first week, outperforming projections and breaking records along the way. That it topped Infinity War's entire offshore run in just 10 days is indeed impressive. The conundrum on a lot of minds is now can-it beat Avatar globally. While it is possible that Endgame is comfortably above $2B worldwide through Sunday, sources are still cautious about calling this MCU pic the all-time champ. It could conceivably need another approximately $600M globally to get there. As we've noted previously, we are in uncharted waters with this incredibly anticipated movie. The film has little in its path over the coming frames, although there is a belief that Detective Pikachu could surprise next session, and certainly a lot will depend on how the Endgame midweeks go next week when markets come back down to Earth out of holiday play and the

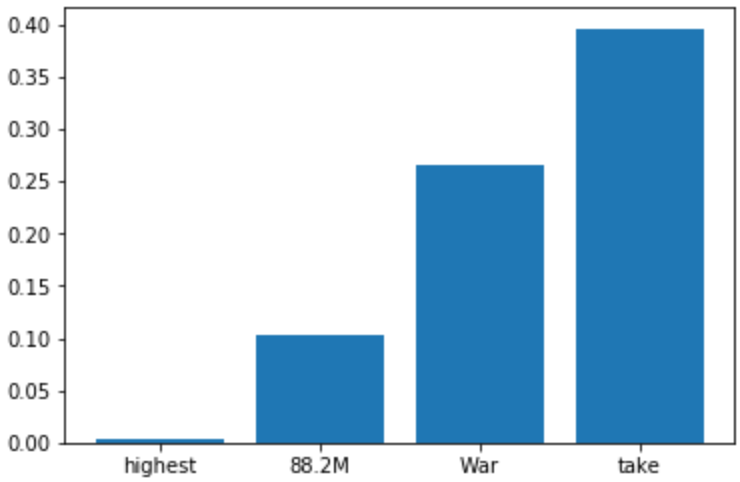
initial rush. Here are the Top 10 markets through Friday: China ($547.4M/does not include Saturday estimates above), UK ($75.6M), Korea ($68.4M), Mexico ($53.5M), Brazil ($47M), India ($45.7M), Germany ($39.9M), Australia ($39.8M), France ($37.3M) and Japan ($29.8M). More to come on Sunday.

Source:

https://deadline.com/2019/05/avengers-endgame-second-weekend-china-global-international-box-office-1202607401/

Prediction: Entertainment

Model Interpretation:



Note that that the graph is not an accurate representation of the model, it is just visualized by getting the activation weights, not the whole gradient of the model itself. But it shows that the word *take* and *war* are the most relevant words in this input. It doesn't also show that if the 2 words are removed, the classification will be different, it just shows that the top 4 words above are the *words which the attention mechanisms focus too.* Also, it is also possible that having 2 layers for this model, and having the attention mechanisms after the 2nd layer, other words may then be filtered through the first layer.

To finalize this project, all objectives were achieved. First, an accuracy rating of above 90% or higher was produced by both models on the English dataset, while an accuracy of 75% and higher was accomplished on the Filipino dataset., thus it is declared that the models indeed accurately classified the news articles. Second, thanks to the utilization of Attention Mechanisms, the authors were able to identify which words were important to classify the inputs.

For the recommendations to those researchers or students who wants to build up with this project, it is derived that there is no trouble for the English dataset, and it classifies human-like performance. However, in the Filipino dataset, it is even not above 90%. One recommendation would be trying 4 and above LSTM layers to reach 90%, or trying their own Word Embeddings for the Filipino Words.

References:

[1]Changshun Du and Lei Huang. 2018. Text Classification Research with Attention-based Recurrent Neural Networks. *International Journal of Computers Communications & Control*13, 1 (2018), 50. DOI:http://dx.doi.org/10.15837/ijccc.2018.1.3142

[2] Philippe Giguere, Francois Laviolette, Gael Letarte, and Frederik Paradis. 2018. Importance of Self-Attention for Sentiment Analysis. *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP (2018).*

[3] Yequan Wang, Minlie Huang, Xiaoyan Zhu, and Li Zhao. 2016. Attention-based LSTM for Aspect-level Sentiment Classification. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (2016). DOI:http://dx.doi.org/10.18653/v1/d16-1058

[4] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In Advances in neural information processing systems (pp. 5998-6008).

[5] Fan Yang, Arjun Mukherjee, and Eduard Dragut. 2017. Satirical News Detection and Analysis using Attention Mechanism and Linguistic Features. *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*(2017). DOI:http://dx.doi.org/10.18653/v1/d17-1211

[6] Anon.Retrieved April 29, 2019 from https://colab.research.google.com/

[7] Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global Vectors for Word Representation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) (2014). DOI:http://dx.doi.org/10.3115/v1/d14-1162

[8] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching Word Vectors with Subword Information. Transactions of the Association for Computational Linguistics 5 (2017), 135–146. DOI:http://dx.doi.org/10.1162/tacl_a_00051

[9] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems (pp. 4765-4774).

[10] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). Why should i trust you?: Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1135-1144). ACM.

[11] Mudrakarta, P. K., Taly, A., Sundararajan, M., & Dhamdhere, K. (2018). Did the model understand the question?. arXiv preprint arXiv:1805.05492.

[12] Sundararajan, M., Taly, A., & Yan, Q. (2017, August). Axiomatic attribution for deep networks. In Proceedings of the 34th International Conference on Machine Learning-Volume 70 (pp. 3319-3328). JMLR. org.

Data Gatherers	**Training Slaves**	**Developers**	**Documentation**
Karl Buchan	Ezekiel David	Justine Clemente	Albert Luz
Edric Roque	Jolo Morales	Julian Yu	Jon Felix Rico
Dale De Ocampo	Emman Franje	Angel Ofalsa	
Danielle Fortuno			
Julius Bautista			
Miguel Katipunan			
Chast Padilla			

Appendix I: Other Test Cases for the English Dataset

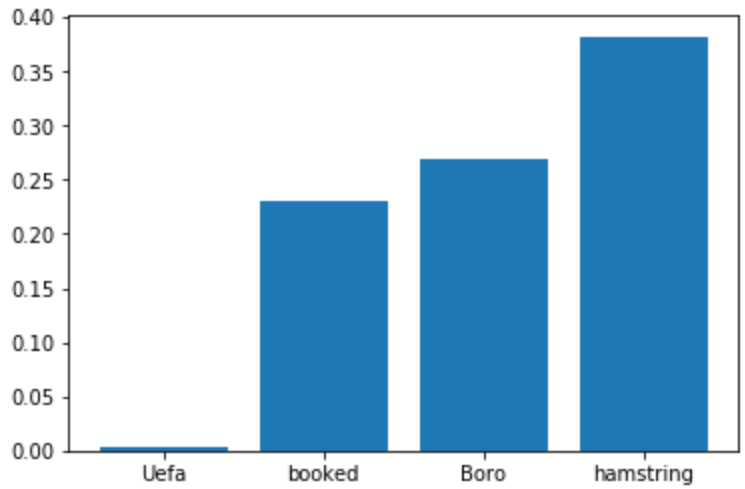
Input Text:

Middlesbrough boss Steve McClaren hopes that Stewart Downing did not seriously damage his hamstring during the Uefa Cup win over Graz AK The winger was carried off on a stretcher in the closing stages as Boro booked their place in the last McClaren said Stewart has tweaked a hamstring which is a blow for us We will assess it in the morning But I m delighted with the match result They say in sport you can win ugly and we certainly did McClaren was relieved to make it through after a tough encounter at The Riverside To be fair to Graz they had a gameplan they knocked the ball forward especially in the first half and we struggled to deal with it In the second half we were better but it was frustrating because we got the goal but one slip and they were back in it Boro face Sporting Lisbon for a place in the quarter finals after the Portuguese side saw off Feyenoord on Thursday And McClaren added Sporting are technically very good Portuguese sides always are But the journey goes on for us and we will now go to their magnificent stadium after the first game here we ll look forward to it Goalscorer Jimmy Floyd Hasselbaink added It wasn t a particularly beautiful match to watch but they made it difficult for us We didn t play well at all But you need a little bit of luck and I think we got that when my shot went through the legs of their goalkeeper I would have liked to play in Holland that would have been a little bit special to me But so is going back to Portugal I ll be playing against some lads I played with at Boavista

True Value: sport

Prediction: sport

{'Uefa': 0.0035266574, 'booked': 0.22970179, 'Boro': 0.268453, 'hamstring': 0.38225123}



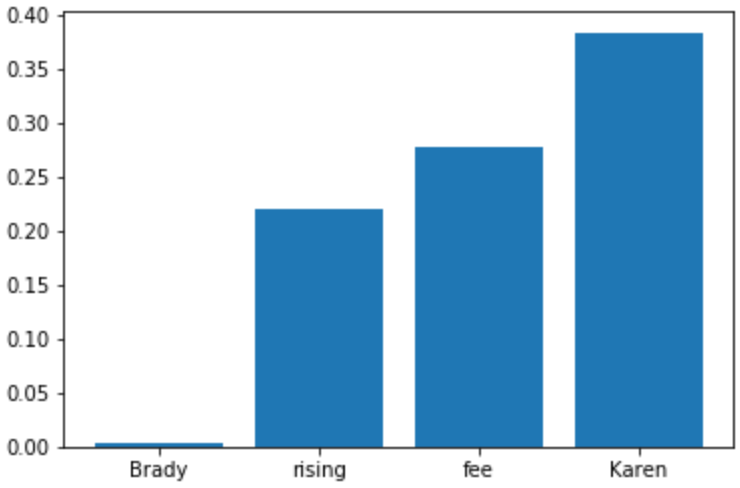
Input Text:

Birmingham have confirmed Blackburn made a bid for Robbie Savage but managing director Karen Brady has called it derisory Rovers have reportedly offered up front for the Wales star with the fee rising to m But Brady told The Sun the bid was a waste of fax paper and my time She added The way things are going all this could affect the relationship between the clubs They ve got into Robbie s head But he s not for sale Savage s

future at Birmingham has been the source of speculation for several weeks with some fans criticising his performances for the club earlier in the season However good displays against West Brom and Aston Villa have impressed Blues fans The crowd gave me a massive standing ovation when I came off on Saturday which was nice he said It was fantastic even though I was criticised by a number of them in recent weeks and on Saturday it showed how much I mean to them It s not for me to say about transfer rumours it s between the two clubs I haven t created the speculation myself I haven t phoned every national newspaper saying Blackburn are trying to buy me It s not up to me Birmingham manager Steve Bruce insists he does not want to sell Savage A lot is said and written about Sav but he has been terrific for Birmingham City the last two and a half years he said The fans love him because he epitomises them He works hard They like people like that and there are not many like him And why the hell should I sell him to someone else I am not interested

True Value: sport

Prediction: sport



Input Text:

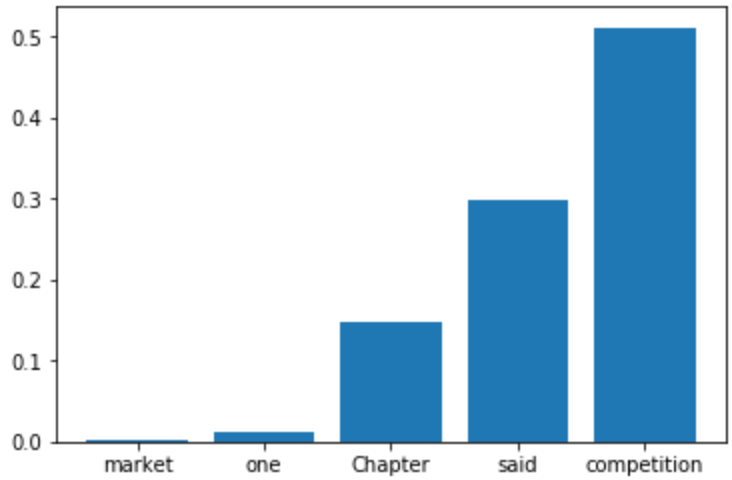
US supermarket group Winn Dixie has filed for bankruptcy protection after succumbing to stiff competition in a market dominated by Wal Mart Winn Dixie once among the most profitable of US grocers said Chapter protection would enable it to successfully restructure It said its stores would remain open but analysts said it would most likely off load a number of sites The Jacksonville Florida based firm has total debts of bn m In its bankruptcy petition it listed its biggest creditor as US foods giant Kraft Foods which it owes m Analysts say Winn Dixie had not kept up with consumers demands and had also been burdened by a number of stores in need of upgrading A month restructuring plan was deemed a failure and following a larger than expected quarterly loss earlier this month Winn Dixie s slide into bankruptcy was widely expected The company s new chief executive Peter Lynch said Winn Dixie would use the Chapter breathing space to take the necessary action to turn itself around This includes achieving significant cost reductions improving the merchandising and customer service in all locations and generating a sense of excitement in the stores he said Yet Evan Mann a senior bond analyst at

Gimme Credit said Mr Lynch s job would not be easy as the bankruptcy would inevitably put off some customers The real big issue is what s going to happen over the next one or two quarters now that they are in bankruptcy and all their customers see this in their local newspapers he said

True Value: business

Prediction: business

{'market': 0.0014284593, 'one': 0.01241009, 'Chapter': 0.14726621, 'said': 0.29805028, 'competition': 0.51109356}



Input Text:

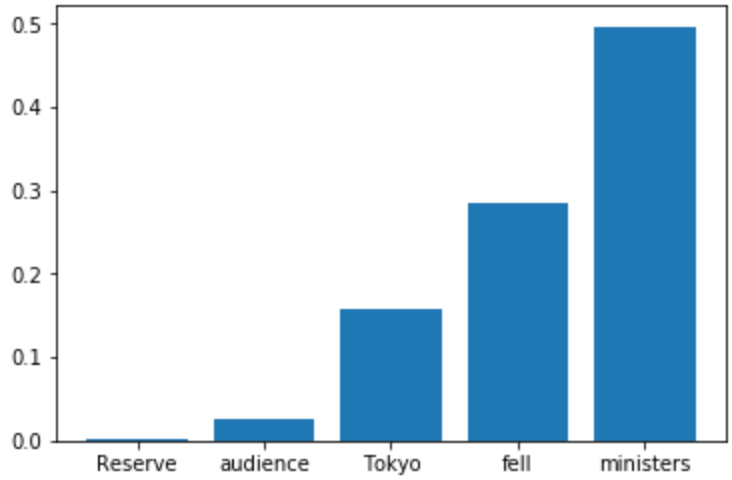
Rising oil prices and the sinking dollar hit shares on Monday after a finance ministers meeting and stern words from Fed chief Alan Greenspan The London FTSE fell while Tokyo s Nikkei dropped its steepest fall in three months G finance ministers said nothing about supporting the dollar whose slide could further jeopardise growth in Japan and Europe And Mr Greenspan warned Asian states could soon stop funding the US deficit On Monday afternoon the euro was close to an all time high against the dollar at above Oil pushed higher too on Monday as investors fretted about cold weather in the US and Europe and a potential output cut from oil producers group Opec although prices had cooled by the end of the day In London the benchmark Brent crude price closed down cents at a barrel while New York light sweet crude closed down cents at a barrel The slide comes as the US has been attempting to talk up the traditional strong dollar policy The latest to pitch in has been President George W Bush himself who told the Asia Pacific Economic Co operation Apec summit in Chile that he remained committed to halving the budget deficit Together with a bn trade gap the red ink spreading across America s public finances is widely seen as a key factor driving the dollar lower And last week US Treasury Secretary John Snow told an audience in the UK that the policy remained unaltered But he also said that the rate was entirely up to the markets a signal which traders took as advice to sell the dollar Some had looked to the G meeting for direction But Mr Snow made clear exchange rates had not been on the agenda For the US government letting the dollar drift is a useful short term fix US exports get more affordable helping perhaps to close the trade gap In the meantime the debt keeps getting bigger with Congress authorising an bn rise in what the US can owe taking the

total to trillion But in a speech on Friday Federal Reserve chairman Alan Greenspan warned that in the longer term things are likely to get tricky At present much of gap in both public debt is covered by selling bonds to Asian states such as Japan and China since the dollar is seen as the world s reserve currency Similarly Asian investment helps bridge the gap in the current account the deficit between what the US as a whole spends and what it earns But already they are turning more cautious an auction of debt in August found few takers And Mr Greenspan said that could turn into a trend if the fall of the dollar kept eating into the value of those investments It seems persuasive that given the size of the US current account deficit a diminished appetite for adding to dollar balances must occur at some point he said

True Value: business

Prediction: business

{'Reserve': 0.0030081873, 'audience': 0.02658897, 'Tokyo': 0.1585051, 'fell': 0.28417423, 'ministers': 0.49648854}



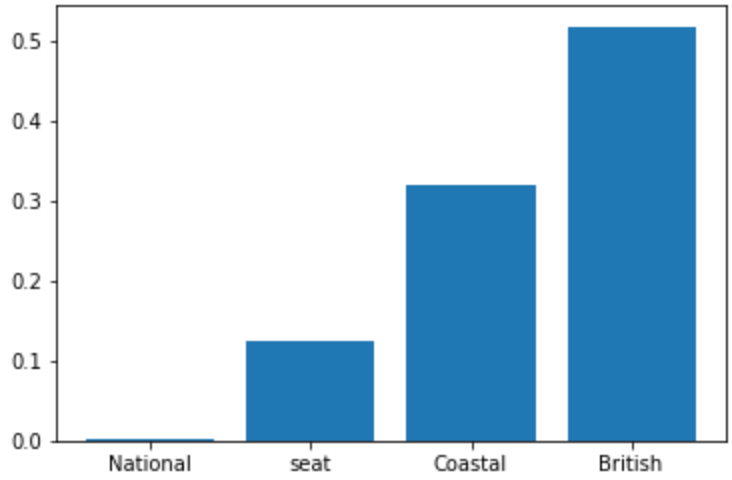
Input Text:

A prospective candidate for the UK Independence Party UKIP has resigned after admitting a brief attachment to the British National Party BNP Nicholas Betts Green who had been selected to fight the Suffolk Coastal seat quit after reports in a newspaper that he attended a BNP meeting The former teacher confirmed he had attended the meeting but said that was the only contact he had with the group Mr Betts Green resigned after being questioned by the party s leadership A UKIP spokesman said Mr Betts Green s resignation followed disclosures in the East Anglian Daily Times last month about his attendance at a BNP meeting He did once attend a BNP meeting He did not like what he saw and heard and will take no further part of it the spokesman added A meeting of Suffolk Coastal UKIP members is due to be held next week to discuss a replacement Mr Betts Green of Woodbridge Suffolk has also resigned as UKIP s branch chairman

True Value: politics

Prediction: politics

{'National': 0.0014631428, 'seat': 0.1244143, 'Coastal': 0.3209951, 'British': 0.5186411}



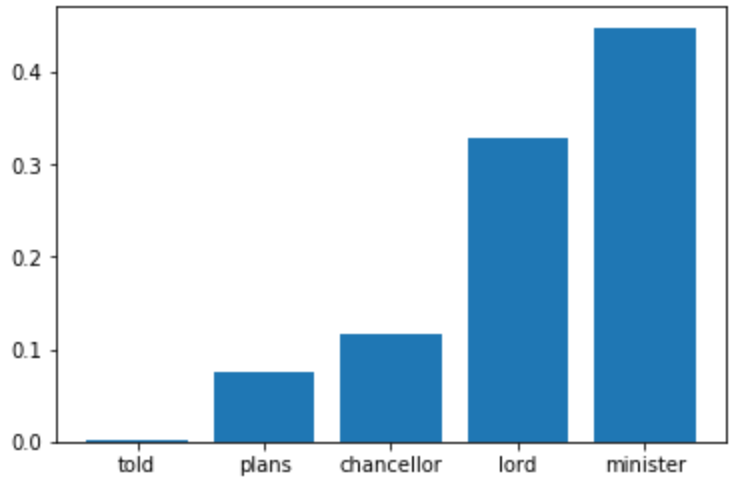
Input Text:

The best person for the job should be appointed lord chancellor and not necessarily a lawyer or MP the courts minister has told MPs Under reforms the post of lord chancellor is to be stripped of its judicial functions The lord chancellor no more needs to be a lawyer than the Secretary of Health needs to be a doctor said courts minister Christopher Leslie The Constitutional Reform Bill was entering its second reading on Monday Mr Leslie said The prime minister should be able to appoint the best person for the job whether they sit in the House of Lords or the House of Commons Under the reforms the Law Lords will also be replaced as the UK s highest legal authority by a Supreme Court and judges will be appointed by an independent panel rather than ministers In December the Lords rejected a plea by current Lord Chancellor Lord Falconer that the holder of the job should not necessarily be a lawyer or a peer The peers voted by to to say in law that lord chancellors must also be peers The debate was carried over from the last Parliamentary session but with an impending general election time is crucial for the government to get the Bill passed Mr Leslie said it was irrelevant whether the post was called Secretary of State for Constitutional Affairs or Lord Chancellor He said What matters most is whether it is reformed so that the post holder no longer has those conflicting duties It is no longer appropriate for a government minister to have such unfettered discretion in the appointment of judges Shadow attorney general Dominic Grieve criticised the government on its plans to change what he said was an exceptional institution providing a champion of the independence of the judiciary The government had initially proposed to take this institution and smash it to pieces Mr Grieve said Convention should be nurtured and celebrated but the government distrusted and disliked it instead He warned that unless ministers backed down over the lord chancellor remaining a member of the House of Lords the government would have great difficulty in getting the Bill through Parliament Former Cabinet minister Douglas Hogg whose father and grandfather served as lord chancellor said the Bill was largely unnecessary bureaucratic and expensive But the Tory MP for Sleaford and North Hykeham admitted the lord chancellor s role and office cannot be frozen in aspic

True Value: politics

Prediction: politics

{'told': 0.0011697117, 'plans': 0.07494365, 'chancellor': 0.11632525, 'lord': 0.32977256, 'minister': 0.44847688}



Input Text:

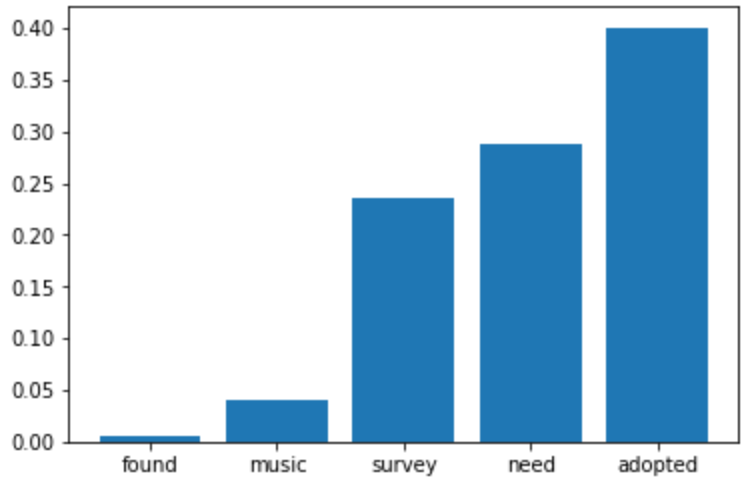
Musicians are embracing the internet as a way of reaching new fans and selling more music a survey has found The study by US researchers Pew Internet suggests musicians do not agree with the tactics adopted by the music industry against file sharing While most considered file sharing as illegal many disagreed with the lawsuits launched against downloaders Even successful artists don t think the lawsuits will benefit musicians said report author Mary Madden For part of the study Pew Internet conducted an online survey of musicians songwriters and music publishers via musician membership organisations between March and April They ranged from full time successful musicians to artists struggling to make a living from their music We looked at more of the independent musicians rather than the rockstars of this industry but that reflects more accurately the state of the music industry Ms Madden told the BBC News website We always hear the views of successful artists like the Britneys of the world but the less successful artists rarely get represented The survey found that musicians were overwhelming positive about the internet rather than seeing it as just a threat to their livelihood Almost all of them used the net for ideas and inspiration with nine out of going online to promote advertise and post their music on the web More than offered free samples online while two thirds sold their music via the net Independent musicians in particular saw the internet as a way to get around the need to land a record contract and reach fans directly Musicians are embracing the internet enthusiastically said Ms Madden They are using the internet to gain inspiration sell it online tracking royalties learning about copyright Perhaps surprisingly opinions about online file sharing were diverse and not as clear cut as those of the record industry Through the Recording Industry Association of America RIAA it has pursued an aggressive campaign through the courts to sue people suspected of sharing copyrighted music But the report suggests this campaign does not have the wholehearted backing of musicians in the US It found that most artists saw file sharing as both good and bad though most agreed that it should be illegal Free downloading has killed opportunities for new bands to break without major funding and

backing said one musician quoted by the report It s hard to keep making records if they don t pay for themselves through sales However said they did not think the lawsuits against song swappers would benefit musicians and songwriters Many suggested that rather than fighting file sharing the music industry needed to recognise the changes it has brought and embrace it Both successful and struggling musicians were more likely to say that the internet has made it possible for them to make more money from their music rather than make it harder for them to protect their material from piracy said Ms Madden

True Value: tech

Prediction: tech

{'found': 0.0062108194, 'music': 0.039815787, 'survey': 0.23595403, 'need': 0.28803468, 'adopted': 0.40044153}



Input Text:

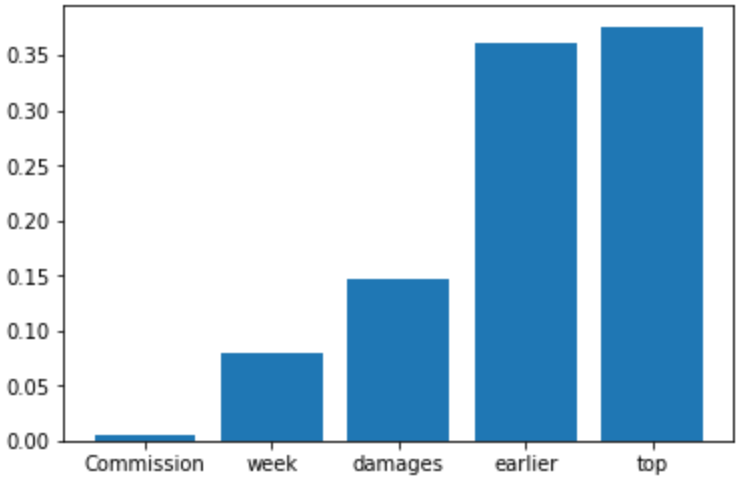
US state Texas has filed a lawsuit against two men believed to be among the world s top five spammers It is seeking millions of dollars in damages in a civil lawsuit filed earlier this week The Texas attorney general said it started the legal action as messages sent by the alleged spammers broke three laws governing e mail marketing The company named in the lawsuit denied any wrongdoing and said it complied with all relevant laws The Texas lawsuit was filed against Ryan Samuel Pitylak a University ofTexas student and Mark Stephen Trotter of California Both are thought to be the top executives in three companies PayPerAction LLC Leadplex LLC and Leadplex Inc that are suspected of sending out many millions of unwanted e mail messages Illegal spam must be stopped said Greg Abbott Texas attorney general announcing the legal action Spam is one of the most aggravating and pervasive problems facing consumers today The attorney general alleges that messages sent by Mr Pitylak and Mr Trotter s companies broke the Controlling the Assault of Non Solicited Pornography and Marketing Act Can Spam as well as the Texas Electronic Mail and Solicitation Act and Texas Deceptive Trade Practices Act All three acts confer cash penalties for each violation of their terms If the men are found guilty and all penalties are applied the two men could face a damages bill running into millions Mr Abbott said the messages sent by the pair broke laws by using misleading subject lines not identifying themselves as adverts and offering services for which they had no

licence to do so in Texas Lawyers for the alleged spammers said the lawsuit was groundless and the two men would defend themselves strongly against the accusations Leadplex and PayPerAction are legitimate internet marketing companies that are in complete compliance with the federal Can Spam Act said Lin Hughes speaking on behalf of Mr Pitylak and Mr Trotter In a similar move the US Federal Trade Commission FTC has won a court order that stops an international group of spammers sending sexually explicit e mail The FTC took the action because the messages being sent violated several parts of the CAN Spam Act In particular the pornographic messages did not identify themselves as being sexually explicit had deceptive subject headings did not have working opt out mechanisms failed to mention they were adverts and did not give the sender s real world address The court order stops the spammers sending e mail and freezes assets prior to a hearing on a permanent injunction

True Value: tech

Prediction: tech

{'Commission': 0.0050971354, 'week': 0.0798604, 'damages': 0.14760503, 'earlier': 0.36114913, 'top': 0.37615058}



Input Text:

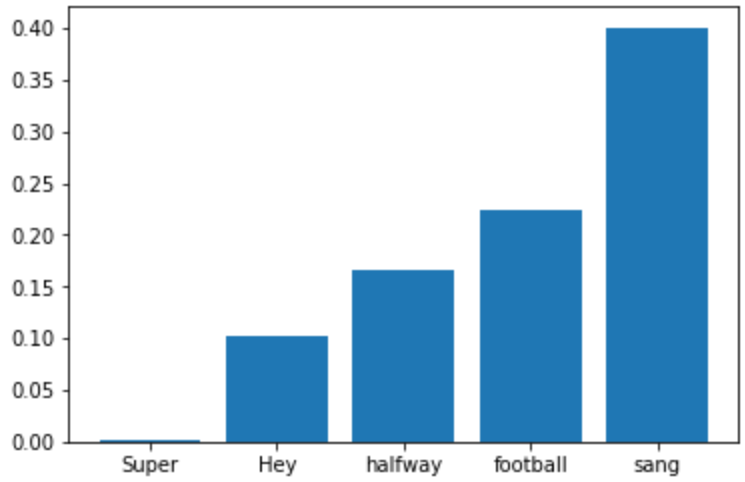
Sir Paul McCartney wowed fans with a live mini concert at American football s Super Bowl and avoided any Janet Jackson style controversies The year old sang Hey Jude and other Beatles songs in a minute set at half time during the game in Florida Last year Jackson exposed a breast during a dance routine causing outrage among millions of TV viewers and landing the CBS TV network a fine Sir Paul however did nothing more racy than remove his jacket as he sang Organisers were widely considered to be playing it safe this year by booking year old Sir Paul for his second Super Bowl show Three years ago he was invited to perform at the first Super Bowl after the September attacks and performed his specially written song Freedom This time he started off the show at the Alltel Stadium in Jacksonville Florida with the Beatles numbers Drive My Car and Get Back He then performed a mellow version of Live And Let Die the James Bond theme he recorded with the band Wings Finally he closed the show with a rousing version of Hey Jude The former Beatle resisted any temptation to refer to Janet Jackson s headline grabbing performance last year instead keeping banter between songs to a minimum in order to squeeze as

much music as he could into his slot The singer removed his black jacket halfway through the show but any fans hoping for a second Nipple gate were to be disappointed as he kept his red sweatshirt on underneath Earlier the Black Eyed Peas and Alicia Keys had provided the night s other high profile entertainment by performing in a pre game show Black Eyed Peas singer Fergie was dressed in a tight orange top and purple hotpants but nothing in her performance was likely to upset TV watchdogs After the controversy last year which saw CBS fined a record by federal regulators Super Bowl organisers had turned to producer Don Mischer to oversee this year s half time show His previous production credits included Olympic opening and closing ceremonies The Super Bowl is watched by an audience of million in the US with many of the people watching are said to tune in specifically to see the entertainment put on around the event Michael Jackson Aerosmith Diana Ross Gloria Estefan and Phil Collins are among the stars who have previously graced the Super Bowl stage

True Value: entertainment

Prediction: entertainment

{'Super': 0.0023704625, 'Hey': 0.101309, 'halfway': 0.16570973, 'football': 0.22432932, 'sang': 0.4006306}



Input Text:

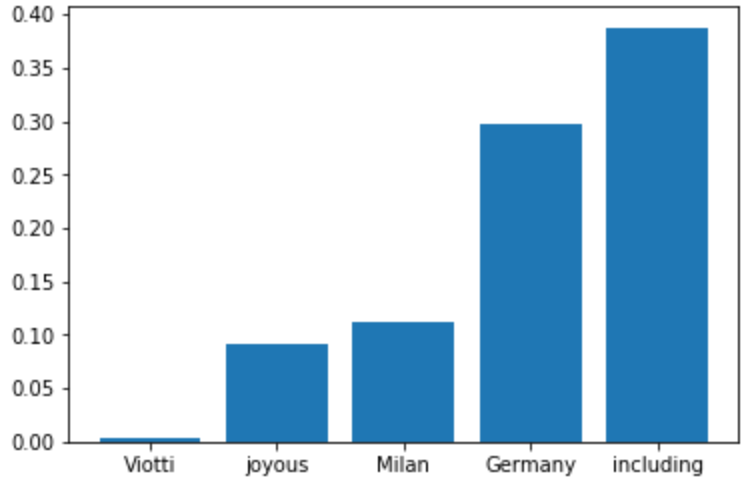
Conductor Marcello Viotti director of Venice s famous La Fenice Theatre has died in Germany at Viotti director of La Fenice since conducted at renowned opera houses worldwide including Milan s La Scala and the Vienna State Opera His time at La Fenice coincided with its reopening in after it was destroyed by fire in He fell into a coma after suffering a stroke during rehearsals for Jules Massenet s Manon last week He conducted some of the best orchestras in the world including the Berlin Philharmonic and the English Chamber Orchestra Viotti was born in Switzerland and studied the piano cello and singing at the Lausanne Conservatory His career breakthrough came in when he won first prize at the Gino Marinuzzi conducting competition in Italy Viotti established himself as chief conductor of the Turin Opera and went on to become chief conductor of Munich s Radio Orchestra At La Fenice Viotti was widely acclaimed for his production of the French composer Massenet s Thais and some of his other productions included Giuseppe Verdi s La Traviata and Richard Strauss s Ariadne auf Naxos The last opera he

directed at La Fenice was Massenet s Le Roi de Lahore Viotti s debut at the New York s Metropolitan Opera came in with Giacomo Puccini s Madame Butterfly followed by La Boheme La Traviata and Fromental Halevy s La Juive Giampaolo Vianello superintendent of the Fenice Theatre Foundation said I am filled with extreme sadness because other than a great artist he is missed as a friend a main character in the latest joyous times during the rebirth of our theatre Viotti s last public performance was on February when he conducted Vincenzo Bellini s Norma at the Vienna State Opera

True Value: entertainment

Prediction: entertainment

{'Viotti': 0.0029512364, 'joyous': 0.09112151, 'Milan': 0.11249195, 'Germany': 0.2974092, 'including': 0.38759613}



Appendix II: Test Cases for the Filipino News Dataset

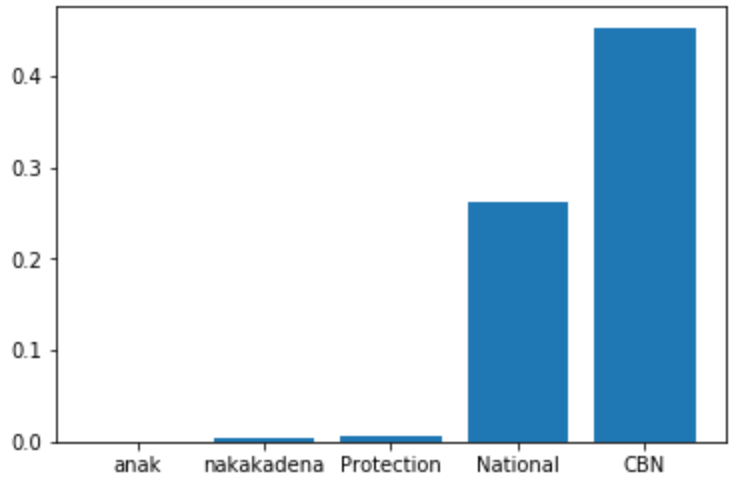
Input Text:

Sa mga larawang ibinigay sa ABS CBN Bicol makikitang naglalaro ang magkapatid na at anyos pero nakakadena ang mga ito Ayon sa nagbigay ng retrato mag iisang buwan na umano itong ginagawa sa mga bata Nang puntahan ng Philippine National Police Women and Children s Protection Desk WCPD ngayong Miyerkoles ang bahay ng mga biktima umamin umano ang ina na kinakadena ang magkakapatid ng kanilang ama Katuwiran ng ina na madalas umanong gumagala ang dalawang bata kaya kinakadena sila pero nakakapaglaro naman sila sa labas ng bahay Nasasangkot din umano sa pagnanakaw ang magkakapatid Dagdag pa ng ina na hindi rin daw siya sigurado kung magbabago pa ang kaniyang mga anak Pero iginiit ng ni Tabaco WCPD chief Police Capt Maria Theresa Berdin labag sa batas ang ginawa ng mga magulang Very obvious naman na hindi tama na ikadena natin ang ating mga anak dahil lang sa rason na sutil or madalas umalis ng bahay dahil alam naman natin na ang kadena hindi akma sa tao aniya Hindi muna kinasuhan ang mga magulang pero regular daw silang bibisitahin ng PNP para hindi maulit ang pagkadena sa mga anak Isasailalim din sa counseling ang pamilya

True Value: Crime

Prediction: Crime

{'anak': 0.00061253493, 'nakakadena': 0.003150984, 'Protection': 0.00591089, 'National': 0.2621863, 'CBN': 0.4529767}



Input Text:

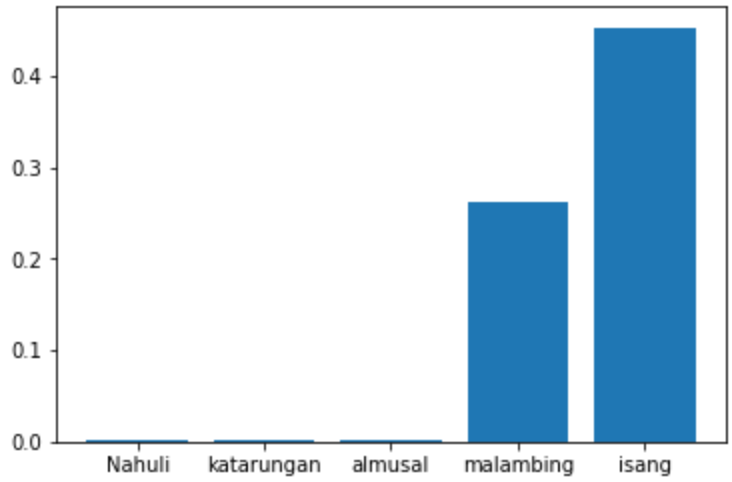
Nahuli sa CCTV ang pamamaril ng isang suspek sa biktimang si alyas Vilma na bumili lang ng almusal noong madaling araw ng Abril Ayon sa kaanak hindi sila makapaniwalang dito hahantong ang buhay ni Vilma na inilarawan niya na mabait at malambing Hiling ng pamilya na mabigyan ng katarungan ang pagkamatay ni Vilma Pangalawa sa limang magkakapatid si Vilma Bagaman tumigil sa pag aaral nais niyang makapagtrabaho para makatulong sa pamilya Palakaibigan din daw ito Dagdag niya walang kalaban laban at posibleng napagkamalan lang ang dalaga Gusto namin na mabigyan ng katarungan Kasi hindi naman din kami matatahimik eh

Makakatulog ba kami nang ganiyan Walang kalaban laban Napagbintangan lang Napagkamalan na iyong hinahanap eh kamukha niya ayon sa kaanak ng biktima Mistaken identity ang lumalabas sa imbestigasyon ng pulisya sa pamamaril Aminado umano ang suspek na napagkamalan ang biktimaItinuturo ang isang alyas Juanabee bilang target ng mga suspek na namaril kay Vilma dahil sa umano y onsehan sa droga Nahuli na ang dalawang suspek pero pinaghahanap pa rin ang isa pa

True Value: Crime

Prediction: Crime

{'Nahuli': 0.0007324239, 'katarungan': 0.0011861521, 'almusal': 0.0015845135, 'malambing': 0.26184458, 'isang': 0.4534792}



Input Text:

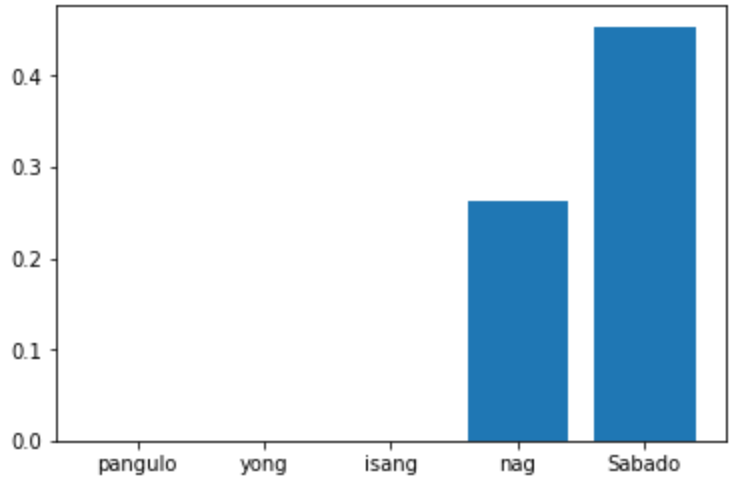
Sa isang talumpati noong Sabado sinabi ni Duterte na may yaman siya dahil sa ina Hoy yong mga dilaw all the time I was with my mother maski nung mayor na ako ang nagpapakain sa kin nanay ko sabi ni Duterte Ang nanay ko nag iwan ng pera sa kin pero kung magkano bakit ko sabihin sa inyo anang pangulo For the longest time I was living with my mother libre lahat pati mura aniya Sa ulat ng PCIJ ipinakita ang Statement of Assets Liabilities and Net Worth SALN ng mga Duterte Sa loob ng taon mula halos nag triple ang yaman ng pangulo na pumalo sa P milyon Higit triple naman ang net worth ni dating Davao City Vice Mayor Paolo Duterte at higit anim na doble ang kay Davao City Mayor Sara Duterte Ito ay kahit mababa ang pasahod sa gobyerno at hindi halos gumalaw ang kita nila sa mga kompanyang kinabibilangan nila Ayon pa sa ulat hindi lahat ng detalye ng ari arian ay nasa SALN ng pangulo Sinagot din ni Duterte ang isa pang ulat ng PCIJ kaugnay ng hindi rehistradong law firm ng anak niya at ilang negosyo ng mga Duterte na wala sa SALN Yong mag advertise kami sa pangalan namin well there s nothing wrong as long as you do not practice and you do not help people get something from government that is illegal ani Duterte Yan ang tingnan nila hindi yang mga law office namin and what happened to our business partnership dagdag niya Binanatan din ng pangulo ang PCIJ Makita mo yang utak ng mga investigative journalism pera pera lang aniya Sa Twitter sumagot ang PCIJ at sinabing karapatan ng

bawat mamamayan na malaman ang SALN ng bawat opisyal ng pamahalaan Iginiit din ng PCIJ na public office is a public trust Nagpadala anila sila ng sulat sa nakaraang buwan kina Pangulong Duterte Sara at Paolo para hingin ang kanilang komento Tiniyak ng pangulo sa publiko na hindi niya pinakikialaman ang pera ng taumbayan Sinabi rin ng pangulo na may masisibak na namang mga opisyal

True Value: Politics

Prediction: Politics

{'pangulo': 0.00032791414, 'yong': 0.00040331122, 'isang': 0.00049087097, 'nag': 0.26189306, 'Sabado': 0.45379528}



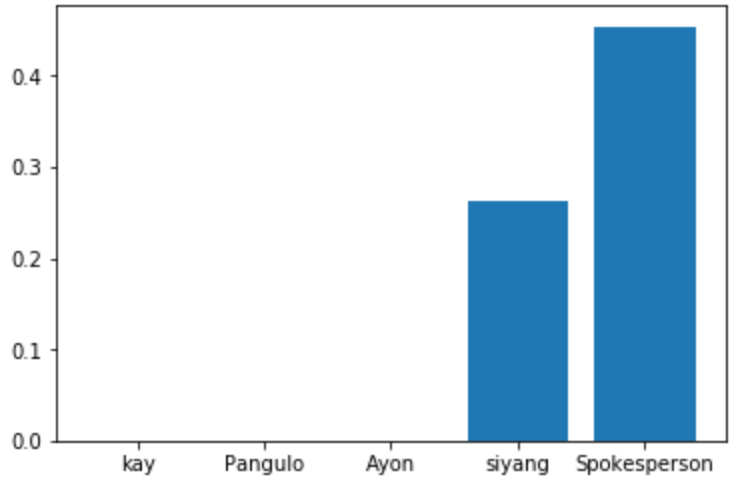
Input Text:

Ayon kay Presidential Spokesperson Salvador Panelo walang babaguhin sa iskedyul ng Pangulo Marami kasi siyang binangga Mga komunista rebelde drug lord kaya natural lang na magkaroon ng threat Itong presidenteng ito wala naman siyang pakialam kahit barilin pa siya sabi ni Panelo sa isang panayam sa DZMM nitong Huwebes Sinabi ang pahayag sa gitna ng pagpapaigting sa seguridad sa mga pinupuntahang pagtitipon ni Duterte Noong Martes pinalibutan ang Pangulo ng isang bullet proof glass habang nagtatalumpati sa isang campaign rally sa Malabon City Samantala sinabi naman ng Philippine National Police PNP na wala pa silang natatanggap na lehitimong banta sa buhay ng Pangulo There is no validated death threat upon the life of Duterte However the PNP will always be vigilant and not let its guard down in providing area and perimeter security whenever he is out for speaking engagements visitations or inspections sabi ni Police Col Bernard Banac tagapagsalita ng pulisya Pagtitiyak pa ni Brig Gen Jose Eriel Niembra group commander ng Presidential Security Group PSG ang mandato nila ay protektahan ang pangulo sa lahat ng oras mayroon man o walang banta sa kaniyang buhay

True Value: Politics

Prediction: Politics

{'kay': 0.00038811698, 'Pangulo': 0.00058411795, 'Ayon': 0.00059104036, 'siyang': 0.26167956, 'Spokesperson': 0.4537521}



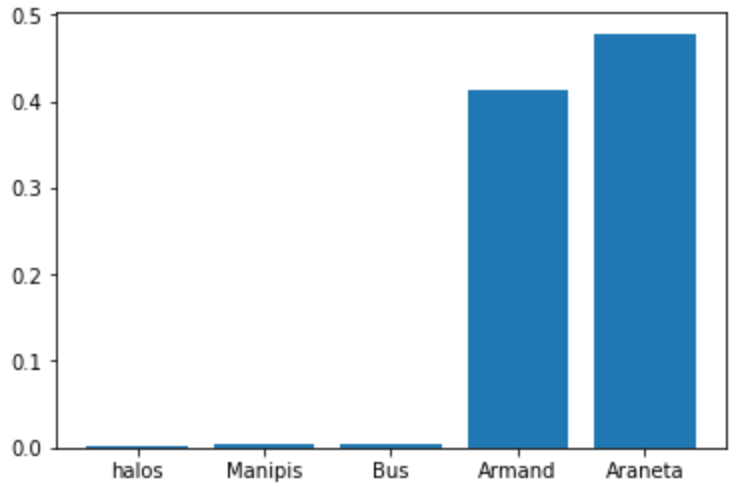
Input Text:

Sa panayam sa DZMM sinabi ni Armand Reyes operations supervisor ng Araneta Bus Station at Bus Port na may nag uuwian na at sinasamantala ang holiday ngayong araw Aniya pumalo na sa halos ang pasahero sa Araneta Bus Station at Bus port hanggang kaninang alas ng umaga Dagdag ni Reyes na aasahan ang pinakamaraming bilang ng pasahero sa Abril Biyernes na huling weekday bago mag Semana Santa Dinamihan na rin nila ang seguridad sa lugar Dinagdagan na po natin ang mga security personnel natin may mga augmentation na po tayo together with mga kapulisan natin dadami rin po yan aniya Manipis pa aniya ang bilang ng mga bumibiyahe kaya inanyayahan na ni Reyes ang ibang pasahero na mag advanced booking at mag online booking para iwas aberya Magdamagan ang operasyon ng Araneta Bus Station kung saan naka terminal ang mga ordinary bus Habang alas ng madaling araw hanggang alas ng gabi naman ang operasyon ng Araneta Bus Port kung saan naka terminal ang mga aircon bus Una nang nagsabi ang Metropolitan Manila Airport Authority at Philippine Coast Guard na naghahanda na sila sa inaasahang dagsa ng pasahero sa Semana Santa mula Abril hanggang

True Value: Vehicles

Prediction: Vehicles

{'halos': 0.0019504543, 'Manipis': 0.004275335, 'Bus': 0.004590821, 'Armand': 0.4132402, 'Araneta': 0.4781894}



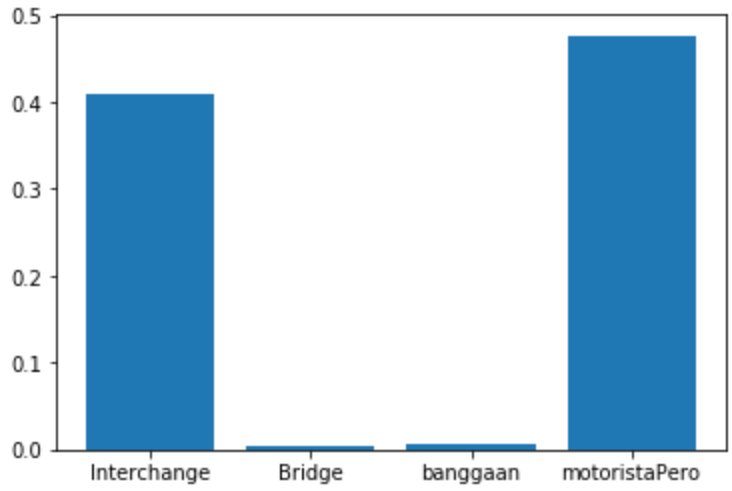
Input Text:

Isinara ang bahagi ng Magallanes Interchange mula Maynila papuntang EDSA northbound sa Cubao Quezon City dahil madalas magkaroon ng banggaan doon ayon sa MMDA Bahagi ng Magallanes Interchange isinara sa mga motoristaPero sa kabila ng pagkaway ng mga MMDA enforcer sa mga motorista para ituro ang isinarang kalsada nalito pa rin ang mga motorista Idinaing din ng mga motorista na mapapalayo ang kanilang biyahe sa pagsara Lalayo po kami ang laking kawalan sa kita namin yon anang taxi driver na si Noel Pagulong Nalito ako akala ko kasi may daan kanina anang motorcycle rider na si Justin Absolig Noong umabot sa higit ang naitalang banggaan sa lugar habang naman noong ALTERNATIBONG RUTAMula Maynila maaaring kumaliwa ang mga motorista papuntang Buendia para makarating sa EDSA Maaari rin mag U turn at dumaan sa Yulo Plaza kumanan sa Pasong Tamo at mag U turn sa ilalim ng Magallanes Interchange Maaari ring dumeretso sa Magallanes Interchange dumaan sa Sales Bridge at mag U turn pabalik sa Magallanes papuntang EDSA Sa mga manggagaling sa Makati at gustong mag U turn ay maaaring dumaan sa U turn sa ilalim ng Magallanes Interchange May mas maikli sanang alternatibong ruta sa Magallanes Village pero hindi ito agad nai coordinate sa mga tauhan ng Barangay Magallanes Nang huling buksan umano ng Magallanes Village ang kalsada nila sa publiko nasira ito Nakiusap ang mga tagaroon na kung bubuksan muli ang kalsada mga light vehicle o kotse lang ang makakadaan Bubuksan daw nila ang kalsada simula gabi ng Lunes hanggang nakasara ang Magallanes Interchange papuntang Cubao

True Value: Vehicles

Prediction: Vehicles

{'Interchange': 0.41043508, 'Bridge': 0.0050932984, 'banggaan': 0.007033387, 'motoristaPero': 0.47676647}



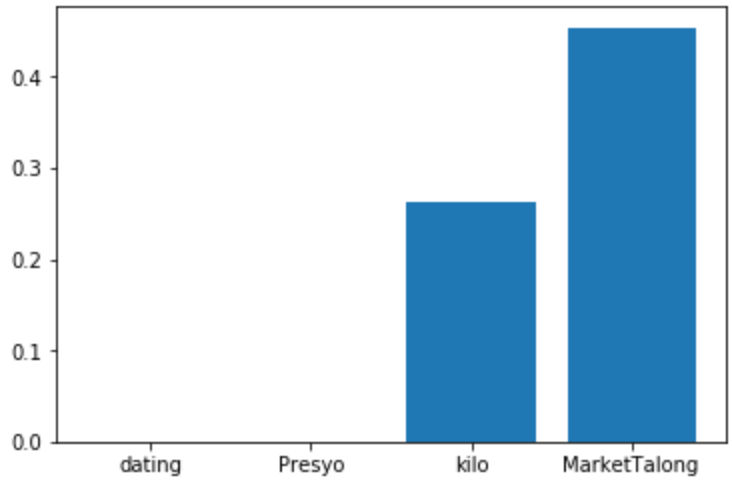
Input Text:

Presyo ng gulay sa Munoz MarketTalong P kilo dating P kilo Carrots P kilo dating P kilo Bawang P kilo dating P kilo Ampalaya P kilo dating P kilo Sayote P kilo dating P kilo Bumaba naman sa P kada kilo ang bell pepper mula P Samantala bumaba ang presyo ng ilang klase ng lokal na bigas sa nasabing palengke Well milled P kilo dating P kilo Regular milled P kilo dating P kilo Tumaas naman ang presyo ng galunggong na P kada kilo ngayon mula P

True Value: Business

Prediction: Business

{'dating': 0.00033481728, 'Presyo': 0.00045312475, 'kilo': 0.26198512, 'MarketTalong': 0.45376945}



Input Text:

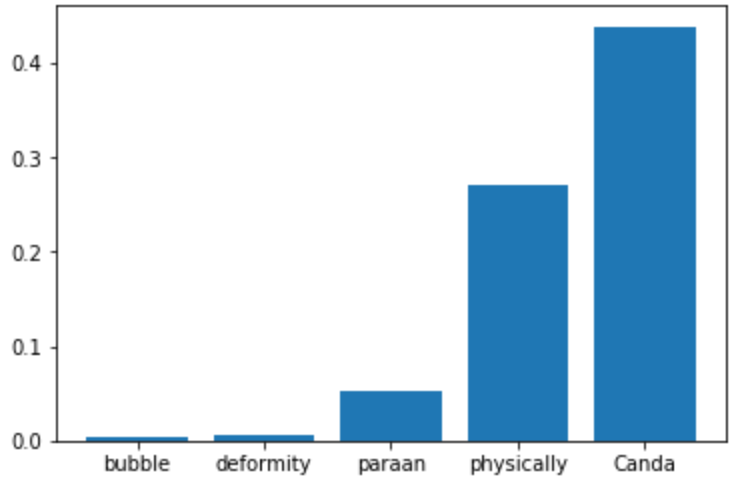
Sa murang edad naranasan ni Canda na tubong Cebu ang labis na pangungutya dahil sa kaniyang deformity I was born with cleft palate so naapektuhan yung pagsasalita ko I ve been bullied since I was a kid Sobrang hirap kasi di ako accepted ng mga tao kasi I m different physically sabi niya Sari saring paraan ang sinubukan ni Canda

upang maging normal sa paningin ng nakararami Nung bata ako kumukuha ako ng bubble gum tapos nilalagay ko sa upper palate para maintindihan yung pagsasalita ko pero hindi pa rin nag work pag alala niya Dahil dito humuhugot na lamang ng lakas sa ina ang binata Payo ng ina just pray and forgive them DANCING IS MY LANGUAGE Dahil sa mga karanasan sa buhay humanap si Canda ng paraan upang maparating ang nais sabihin nang hindi gumagamit ng salita Ang ginawa ko naghanap ako ng talent na hindi kailangan magsalita aniya Dito niya napagtanto ang isang talento na babago sa takbo ng kaniyang buhay Dancing talaga is my language Mula nang mahilig sa pagsayaw pinangarap na ni Canda ang mapanood sa telebisyon partikular na sa programang It s Showtime ng ABS CBN I ve been an avid fan po talaga ng ABS CBN Idol ko po talaga si Billy Crawford ayon sa dancer na makailang beses sumali sa Magpasikat ng It s Showtime at sa Pilipinas Got Talent pero hindi pinapalad Pero nabuhayan ang binata ng pag asa nang magpatawag ng audition ang noo y ipapalabas pa lamang na patimpalak na World of Dance Philippines Nung nakita ko yung World of Dance Philippines sinabi ko talaga na hindi ako magdadalawang isip sumali DREAM COME TRUEDahil sa determinasyon nakapasok sa World of Dance Philippines ang grupo ni Canda na The Exporters at natupad ang kaniyang pangarap na makasayaw sa telebisyon Gayunpaman may kurot sa puso ng binata ang tinamasang tagumpay dahil hindi na niya kapiling ang inang sumuporta sa kaniya mula umpisa Sobang sayang kasi hindi niya napanood yung dream come true ko na makasayaw sa TV sabi ni Canda Pumanaw ang ina niya noong isang taon dahil sa stage colon cancer Pati ang kaniyang pagtatapos sa kursong civil engineering ay hindi na rin natunghayan ng ina Sa kabila nito masaya pa rin ang dancer dahil matapos mag viral ang kanilang audition ay lumapit ang organisasyong Operation Smile para magbigay ng tulong Willing po silang ayusin po yung cleft palate ko banggit ni Ryan We are happy na through your ABS CBN program na discover nila ang isang Ryan who is showing aspiration of the people sabi naman ni Hector Santos ng Operation Smile Sa ngayon abala ang binata sa pagrerebyu bago ang nakatakda niyang board exams sa Mayo para maging ganap na civil engineer Hindi ko talaga makakalimutan yung ABS CBN kasi ang daming naitulong sa akin

True Value: Entertainment

Prediction: Entertainment

{'bubble': 0.004565192, 'deformity': 0.0049736956, 'paraan': 0.053615484, 'physically': 0.269782, 'Canda': 0.438861}


0.4
0.3
0.2
0.1
0.0
bubble
deformity
paraan
physically
Canda