Big Social Data Analytics for Public Health: Predicting Facebook Post Performance using Artificial Neural Networks and Deep Learning

Nadiya Straton[1], Raghava Rao Mukkamala[1], Ravi Vatrapu[1,2]
[1]Centre for Business Data Analytics, Copenhagen Business School, Denmark
[2]Westerdals Oslo School of Arts, Comm & Tech, Norway
{nst.itm, rrm.itm, rv.itm}@cbs.dk

Abstract—**Facebook "post popularity" analysis is fundamental for differentiating between relevant posts and posts with low user engagement and consequently their characteristics. This research study aims at health and care organizations to improve information dissemination on social media platforms by reducing clutter and noise. At the same time, it will help users navigate through vast amount of information in direction of the relevant health and care content. Furthermore, study explores prediction of popularity of healthcare posts on the largest social media platform Facebook. Methodology is presented in this paper to predict user engagement based on eleven characteristics of the post: *Post Type, Hour Span, Facebook Wall Category, Level, Country, isHoliday, Season, Created Year, Month, Day of the Week, Time of the Day*. Finally, post performance prediction is conducted using Artificial Neural Networks (ANN) and Deep Neural Networks (DNN). Different network topology measures are used to achieve best accuracy prediction followed by examples and discussion on why DNN might not be optimal technique for the given data set.**

Index Terms—**Post Performance, Artificial Neural Network (ANN), Deep Neural Network (DNN), Negative Entropy, Purity.**

I. Introduction

Innovative advances in participatory internet make social media platforms such as Facebook an inescapable platform for health care promotion and education [1]. Benefits of using social media platform such as Facebook for public health care information dissemination include expansive reach, interactivity that enables both anonymity and social networking according to personal preferences, relatively low costs to spread health care information compared to traditional media such as newspapers, TV and radio. The above benefits of social media usage have resulted in an information deluge, as individuals and organizations upload 350 million of photos to Facebook per day and generate 4 million likes every minute [2]. Thus, majority of posts on social media go unnoticed by target users or even worse, inaccurate or misleading information can go viral. For example, in our data set, out of all the posts submitted to 153 Facebook health care walls, from 2006 to 2015 only top 5% received '296' or more post likes, '43' or more - post share and '19' or more comments. Only handful are highly popular posts, gathering more than 100,000 likes. 25% of health care posts did not receive a single Post Like or Share. Bottom 50% of health care posts received '3' Post Likes at the maximum and as such almost half of the posted data is a wasted effort.

One explanation for the fact that a majority of the Facebook posts receive little user attention, reason could be that well-known or well-resourced companies can afford to buy more exposure than their smaller counterparts. Another explanation could be post characteristics that impact post performance. This paper aims to discover the existence of such post characteristics and measure their influence on post engagement performance. Research question, objective and propositions are listed below:

A. Research Question

What, if any, are the characteristic dimensions of Facebook posts that account for post performance in the health and care domains?

B. Research Objective

The objective of this paper is to find the right predictive model that can help health care organisations in terms of their social media marketing strategy and tactics. We test different algorithms, models and statistical approaches to find the most effective method to evaluate post performance and then make predictions based on the number of relevant attributes with Artificial Neural Networks (ANN) and Deep Neural Networks (DNN).

C. Research Propositions

1) Simple models can lead to accurate and reliable predictions of Facebook post performance.
2) Selecting the right features and understanding health care domain-specific aspects of data leads to better results.
3) Deep learning can improve data analyses and achieve higher accuracy results than simple ANN network due to the increased number of hidden units and network layers that disseminate data points by weighing them in each layer.

II. Related Work

Extant literature in Big Social Data Analytics is dominated by research on Twitter [3]. In the public health care domain, current state-of-the-art is characterised by a focus on application of statistics and machine learning to textual or semi-structured data primarily from Twitter [4] with recent empirical research on Facebook data sets [5][6].

A. Supervised Learning in Big Social Data Analytics

Not all of the algorithms applied on big social data sets are computationally as expensive as Deep Neural Network. Previous work [4] shows that classification algorithms achieve higher accuracy prediction rates with textual content and lower accuracy rates with quantitative/categorical content. In this paper, ANN is applied on a big social data set from Facebook to forecast certain attributes and their likelihood of belonging to one or the other of the popularity clusters. It has been suggested that the prediction accuracy of how popular the content will be in the future might depend on the right choice of the model [7].

B. Artificial Neural Networks (ANN)

ANN has previously been successfully used on text data and [8] employs ANN for sentiment classification on Twitter. More specifically, [8] use ANN with n-gram analyses for feature extraction. Authors developed DAN2 (a Dynamic Architecture for Artificial Neural Networks) using a feed forward approach with input, hidden and output layer. However, number of hidden layers in not fixed a priori as in the current research. Instead, layers are sequentially and dynamically generated through knowledge propagation, adjusting it forward to the next layer, until the desired level of network performance criteria is reached [8]. [8] credit favourable performance of the DAN2 to the fact that network is trained using all observations in the training set simultaneously, so as to minimize a mean squared error (MSE) value. Moreover, their approach evidenced better recall score in comparison to SVM performance. On the other hand, [9] applied ANN to make predictions based on quantitative or categorical data instead of text. Their approach is very close to traditional approach applied in the current research to quantitative and categorical data from Facebook and employs classification to predict expected revenue range from the box office sales before actual movie release. Then researchers compare ANN results to prediction with statistical models and find ANN to achieve highest accuracy of around 37% in comparison to Logistic Regression, Discriminant Analyses and C&RT [9]. Similarly to current research [10] mentions [9] to have treated the prediction problem as a classification problem that classifies movies into 9 categories (output units) and makes predictions over actual numbers with two hidden unit network [10]. [10] suggests that prediction over actual numbers is the reason on why authors achieve fairly low accuracy results [10].

Accuracy of 37% is much lower than the one presented in this paper: the best prediction with ANN on health care data from Facebook achieved 69% accuracy rate, with two hidden layers and five hidden units in each layer.

C. Deep Learning

Deep neural networks have been successfully applied to image and voice recognition, whether it is social media data from Youtube or known data sets with images such as MINST (handwritten digits). Extant literature suggests that the combination of the right model and high computational power usually lead to good results. There are different architectures used with Deep Neural Networks (DNN) such as convolutional neural networks [11][12][13], deep belief networks [14] or deep restricted Boltzmann machines [15] among others [16]. [11] used GPU to improve the run time, regularization to avoid over-fit and achieved very good error rates on a challenging data set with deep neural network: 37.5% and 17% error rates on 1.2 million image classification with 5 convolutional layers and showed that removal of any of the middle layers would result in 2% performance loss. In this paper, we use neither regularisation method to avoid over-fit nor GPU implementation to reduce run-time problem with increased number of layers. [12] used convolutional network with 7 layers and softmax activation and obtained slightly better results than [11] with 14.8% prediction accuracy. Removing middle layers from the network lead to decrease in the performance. Therefore, [12] suggest that depth and size of the layers is important for obtaining good performance. [13] showed application of (3×3) convolutional neural network to classify images by pushing depth gradually to $16 - 19$ weight layers. With test error rate of 7.3% authors demonstrated the importance of depth in the visual representation. [17] applied convolutional neural network (CNN) with 22 layers and filter size (5x5). Training was conducted with asynchronous stochastic gradient descent that fixed learning rate schedule. They achieved 6.7% lowest top 5 classification error rate on both validation and testing and credit success to optimal sparse structure by readily available dense building blocks. It is also possible to achieve good results with simpler architectures as presented by [18] involving deep multi-layer perceptron (MLP) with back propagation that yields 0.35% error rate on MINST hand-written digits. However, network with more layers does not always have a better structure as [19] have empirically demonstrated that shallow feed-forward nets can learn complex functions previously learned by deep nets and achieve the same accuracies, also by using the same number of parameters. In this paper, simpler network topology will be compared to more complex deeper networks to find if DNN contributes to higher prediction accuracy.

D. Model/Algorithm Validation

Previous studies have evaluated post popularity and model accuracy. For example, [20] use RMSE (Root Mean Squared Error) and Kendall coefficient while [21] computed precision and recall values based on authors aggregate survey ratings. [22] clustered social media documents with incremental clustering method and used combined NMI and B-Cubed scores on the validation set to determine the weight of each cluster and to find out how much information is shared between actual 'ground truth' events and each associated clustering assignment.

III. Data Set Description and Research Methodology

Descriptive statistics will initially be used to visualize data and perform reductions if necessary. Unsupervised learning techniques are applied to achieve post engagement attributes, then suitable performance measures are used to ascertain the quality of the data analysis. In order to predict if post will perform well or not supervised learning techniques are used on labelled data derived from clustering results.

A. Data set description and process flow

Start date: **2006-01-01** End date: **2015-12-30**		
Number of Facebook Walls: **153**		
Activity	**No. of Actions**	**Unique Actors**
Facebook Page Likes	10, 476, 523	–
Facebook Posts	280, 534	101, 351
Post Shares	4, 225, 739	–
Likes on Posts	24, 331, 261	7, 129, 957
Comments	1, 734, 154	788, 297
Likes on Comments	1, 507, 687	493, 266
Comment Replies	208, 512	100, 379
Likes on Comment Replies	176, 920	88, 202
Total	42, 941, 330	7, 531, 865[1]

Table I: Overall Statistics of Public Health Facebook Dataset

Data from 153 public Facebook walls of various public health organizations was collected using Social Data Analytic Tool (SODATO) [23]. These walls include national as well as international agencies, organisations as well as individual bloggers. The total dataset contains information about around 43 million Facebook actions that happened during a time period of 10 years as shown in table I. Majority of actions are *Likes on Posts* (around 55%) and the dataset contains around 280 000 Facebook posts. Around 34% of dataset are *Facebook Page Likes* and *Post Shares* and Facebook does not provide user information in respect of these items. In the entire dataset, there are around 7.5 million unique users and as one can notice from table I, the prominent action performed by the users is *like*.

The overall methodology for the research work is displayed in the figure 1.

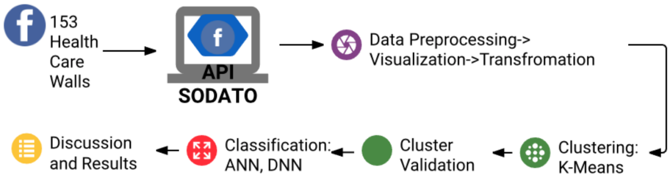


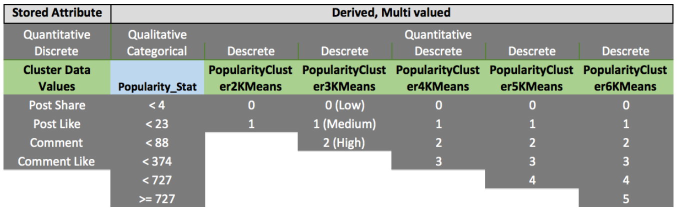
Figure 1: Data Process Flow

B. Post Performance Frame

In order to measure post performance and post popularity using clustering algorithms, we have chosen the four most related/correlated attributes that show post performance: *Post Like*, *Post Share*, *Comment*, and *Comment Like*. Pearson correlation and coefficient of determination (r^2) [24] were used to frame engagement and indicate linear association between attributes. [25] interpret coefficient, as proportion of fluctuation of one variable that is predictable from the other variable, variance 'explained' by the regression model is useful as a measure of success when predicting a dependent variable from independent variables. Moreover, coefficient is asymptotically independent of the sample size n [25].

C. K-means

There is no previous study on using K-means and GMM clustering with health care data from Facebook, therefore it was necessary to use performance measures to evaluate cluster quality against labelled data. Data structure derived as a result of manual classification and clustering with five different scenarios: 2,3,4,5,6 cluster assignments is presented in Figure 2. Performance measures: Negative Entropy, Purity, Rand, Jaccard, Completeness, Homogeneity, Mutual Info,V-Measure, Adjusted Rand show how well K-means clustered data in comparison to manually assigned labels. Rand depends on number of clusters and size of the data set and therefore might vary with the different set up, as suggested in [26]. Completeness, Homogeneity and V-measure can be applied to any clustering solution, as are independent of the number of clusters, size of the data set and algorithm. Rand counts correctly classified pairs of elements and ranges from 0 to 1, with 1 being correctly classified and 0 - misclassified. Adjusted Rand has the same range from 0 (independent cluster results) to 1 (identical cluster results) and is drawn with the fixed number of elements in each cluster as mentioned in [26]. Jaccard performs similar measurement to Rand, however disregards the pair of elements that are in different clusters.



Stored Attribute	**Derived, Multi valued**					
Quantitative Discrete	Qualitative Categorical	Quantitative				
		Descrete	Descrete	Descrete	Descrete	Descrete
Cluster Data Values	**Popularity_Stat**	**PopularityCluster2KMeans**	**PopularityCluster3KMeans**	**PopularityCluster4KMeans**	**PopularityCluster5KMeans**	**PopularityCluster6KMeans**
Post Share	< 4	0	0 (Low)	0	0	0
Post Like	< 23	1	1 (Medium)	1	1	1
Comment	< 88		2 (High)	2	2	2
Comment Like	< 374			3	3	3
	< 727				4	4
	>= 727					5

Figure 2: Data Structure: clustered and manually classified data.

$$J(C, C') = \frac{n_{11}}{n_{11} + n_{10} + n_{01}}$$

V-Measure is defined as the harmonic mean of homogeneity and completeness. "V-Measure evaluates the quality of clustering not a post-hoc class-cluster mapping", [27].

$$V_{ß} = \frac{(1 + ß) * h * c}{(ß * h) + c}$$

where h - homogeneity, c - completeness.

"Homogeneity is maximized, when class distribution within each cluster is totally skewed to a single class, that is, zero entropy" as mentioned in [27].

$$h = \begin{cases} 1 & if H(C, K) = 0 \\ 1 \quad - \frac{H(C|K)}{H(C,K)} & else \end{cases} \Bigg\}$$

where C is a set of classes, and K is a set of clusters.

[1] Total unique actors for the whole dataset

Golden standard of Completeness score according to [27] is to put all samples of the same class into one cluster:

$$c = \left\{\begin{matrix} 1 & & if H(K,C) = 0 \\ 1 & -\frac{H(K|C)}{H(K,C)} & else \end{matrix}\right\}$$

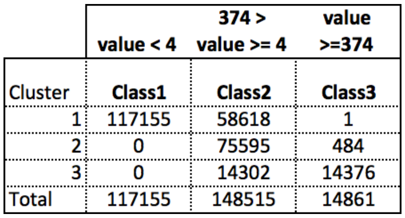
[27] suggests that Purity and Entropy are likely to improve with increase in the number of clusters and disregards completeness criterion in its calculation, therefore is not ideal measurement.

$$Purity = \sum_{r=1}^{k} \frac{1}{n} \underset{i}{max}(n_r^i)$$

$$Entropy = \sum_{r=1}^{k} \frac{n_r}{n}(-\frac{1}{log_q} \sum_{i=1}^{q} \frac{n_r^i}{n_r} log \frac{n_r^i}{n_r})$$

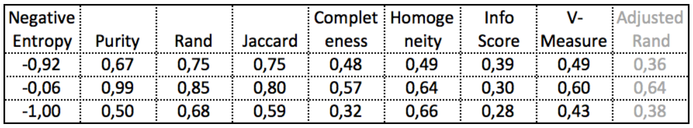
where q is the number of classes, k is the number of clusters, n_r is the size of cluster r and n_r^i is the number of data points in class i clustered in the cluster r.

Each Performance measure has their advantages and disadvantages, therefore to make more precise and fair evaluation of the cluster results, all the above mentioned measures were applied. The best *Negative Entropy, Purity, Rand, Jaccard, Completeness, Homogeneity, V-Measure* scores were observed in 2 and 3 cluster set up in contrast to 6, 5 and 4. Contradictory finding to the statement given by [27] state that: ”Purity and Entropy are likely to improve with increase in the number of clusters”. Tables II and III show 3 clusters (clustered with K-means algorithm) vs. manually assigned classes and their performance measures: Negative Entropy, Purity, Rand, Jaccard, Completeness, Homogeneity Mutual Info, V-Measure Score and Adjusted Rand. Cluster performance measures were instrumental in evaluating clustering results and deciding on the number of clusters.



	value < 4	374 > value >= 4	value >=374
Cluster	Class1	Class2	Class3
1	117155	58618	1
2	0	75595	484
3	0	14302	14376
Total	117155	148515	14861

Table II: 3 Clusters vs Classes



Negative Entropy	Purity	Rand	Jaccard	Completeness	Homogeneity	Info Score	V-Measure	Adjusted Rand
-0,92	0,67	0,75	0,75	0,48	0,49	0,39	0,49	0,36
-0,06	0,99	0,85	0,80	0,57	0,64	0,30	0,60	0,64
-1,00	0,50	0,68	0,59	0,32	0,66	0,28	0,43	0,38

Table III: Performance Measures

D. Post Performance Prediction with ANN and DNN

Data set analysed in this section has combination of quantitative and qualitative attributes. Some of the attributes were part of the data from the beginning, some were derived to achieve better insight into features and some attributes were disregarded at this point, as were not relevant for the analyses or already served a purpose for other derived attributes.

Results from K-Means clustering are discrete values from 0 to 2 and represent ’Low’, ’Medium’ and ’High engagement’ clusters. These clusters will represent dependent output parameters. ANN and DNN will predict engagement classes over independent input parameters. They are represented by quantitative and qualitative attributes. Quantitative: *isHoliday, Season, Created Year, Month, Day of Week, Time of Day, Hour Span between Create and Update date* and are discrete. Qualitative: *Post Type, Facebook Wall Category, Level* and *Country*. In its turn each attribute includes features used for analyses in classification section of the report and feature mining. *isHoliday* attribute aims to answer if engagement activity falls/rises during holiday season. Similarly *Day of Week* and *Time of Day* might highlight more favorable slot to post a message. *Hour span* helps to find out if engagement activity was re-posted/updated and how big is the gap between create and update date. Artificial Neural Networks (ANN) will be applied to estimate which attributes contribute the most to the prediction results. Additionally, classification matrix will show if performance of the model is better than simply predicting all outputs to be the largest class in the training data set.

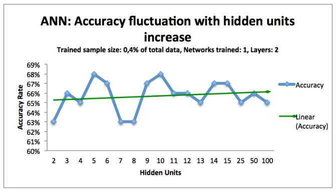
Network topology was established through number of hidden layers, number of nodes and activation function. Definition of neural network according to Pang et al.: ”Single layer neural network is a perceptron that performs complex operations in one layer/can create one hyperplane and therefore cannot find optimal solution, as opposed to multi-layered perceptron. Latter consists of a number of hidden nodes that can be considered as perceptrons located at the layers, while the output layer simply combines the results of the perceptrons to the decision boundary” [28]. Therefore this research applies multi-layered feed froward neural network to find optimal solution and will perform several reductions to choose the right model.

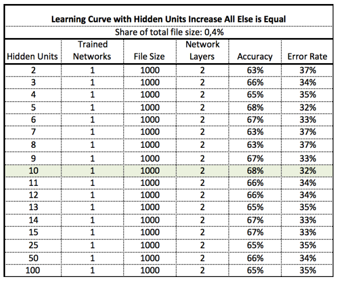
”The goal of ANN learning algorithm is to determine a set of weights that minimize the sum of squared errors” according to Pang et al. [28]:

$$E(w) = \frac{1}{2} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

”Sum squared errors depends on w because the predicted class $\hat{y}_i$ is the function of the weights assigned to the hidden and output nodes.” [28]. Since numerical values of ’independent variables’ have various scales, they were standardized.

1) Artificial Neural Network Structure: Optimal neural network topology was selected to achieve best possible results. Following table shows results of 17 algorithm runs with varied, increasing number of hidden units from 2 to 100, fixed network layers - 2 and one trained network with each iteration. Figure and Table IV show accuracy fluctuations with increase in the number of hidden units. File size, number of layers (2) and trained networks (1) are kept constant.

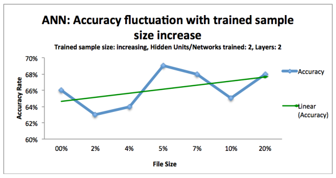


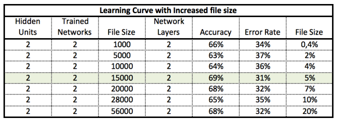


Learning Curve with Hidden Units Increase All Else is Equal					
Share of total file size: 0,4%					
Hidden Units	Trained Networks	File Size	Network Layers	Accuracy	Error Rate
2	1	1000	2	63%	37%
3	1	1000	2	66%	34%
4	1	1000	2	65%	35%
5	1	1000	2	68%	32%
6	1	1000	2	67%	33%
7	1	1000	2	63%	37%
8	1	1000	2	63%	37%
9	1	1000	2	67%	33%
10	1	1000	2	68%	32%
11	1	1000	2	66%	34%
12	1	1000	2	66%	34%
13	1	1000	2	65%	35%
14	1	1000	2	67%	33%
15	1	1000	2	67%	33%
25	1	1000	2	65%	35%
50	1	1000	2	66%	34%
100	1	1000	2	65%	35%

Table IV: Hidden Units Increase, file size: 1000 data points, 0.4% train sample

Overall linear development of hidden units and accuracy rate shows slightly upward trend. However, with more than ten hidden units in each of the layers, accuracy rate decreases. Increase in the number of units does not contribute to the quality of the training results. The best accuracy of 69% is achieved with 15 hidden units and trained file size of 5%. General trend shows linear growth between hidden units/trained networks and prediction accuracy. Increase in the training sample size lead to 1% increase in accuracy from 68% to 69% with corresponding file sizes increase from 0.4% to 10%. However there is also considerable overload on the resources. Furthermore, it was interesting to research if increase in the file size from 0.4% to 20% leads to higher accuracy than 69%. Hidden units/trained networks and number of layers were kept constant. Figure and Table V show accuracy fluctuations with increased trained same size. Number of hidden units, trained networks and number of layers (2) are kept constant.





Learning Curve with Increased file size						
Hidden Units	Trained Networks	File Size	Network Layers	Accuracy	Error Rate	File Size
2	2	1000	2	66%	34%	0,4%
2	2	5000	2	63%	37%	2%
2	2	10000	2	64%	36%	4%
2	2	15000	2	69%	31%	5%
2	2	20000	2	68%	32%	7%
2	2	28000	2	65%	35%	10%
2	2	56000	2	68%	32%	20%

Table V: File size increase from 0.5 to 20%

There is a positive linear relation between accuracy rates and file size increase. The best rate of 69% is reached with 2 hidden layers and 2 hidden units, trained sample size of 5% (15000 lines). Figure 3 shows such Neural Network with eight input parameters, 2 layers and 2 hidden units in each of the layers and 3 output nodes that correspond to each of the engagement clusters. The bigger the file size, the less hidden units suffice to achieve higher accuracy. However, might be pre-mature to make concrete conclusion, as only up to 20% of the trained sample size was explored. Moreover, accuracy rate improvement is rather small, from 66% to 69% with file size increase from 0.4% to 5% and then 20%. Since data is picked randomly from a sample, results can be quite diverse each time. File size might not be the sole contributor to accuracy of the results, but rather weights chosen by Neural Network algorithm with each network training, number of layers and number of hidden units at each layer and their combination. Best accuracy result of 69% is slightly low when comparing to results in the literature achieved with text and image data. Reason can be due to data sparsity and explanation given by [10], when authors mention work of [9] and suggest that prediction over actual numbers (categorical data in the current research) is the reason of lower accuracy results II-B.

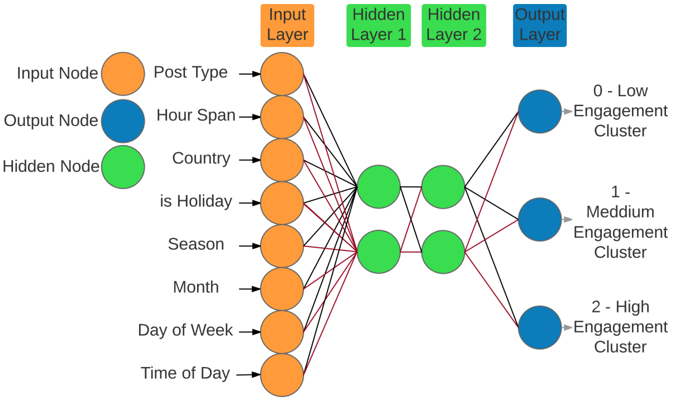


Figure 3: Neural Net, 2 hidden layers, 2 hidden units

Evidence in this section shows that right neural network architecture can be important for achieving more accurate results. Figure 4 shows multi-layered feed forward neural network with 5 hidden units and 2 layers. Network architecture from Figure 4 is used in the result section of the paper. Data points from 8 attributes: *Post Type, Hour Span, Time Of Day, Day Of Week, Month, Season, IsHoliday, Country Code* are classified through two layers with five hidden units in each. Weights are re-calculated with each neural network initialization, in this case network is initiated 5 times with the best outcome. Nodes in one layer are only connected to the nodes of the next layer.

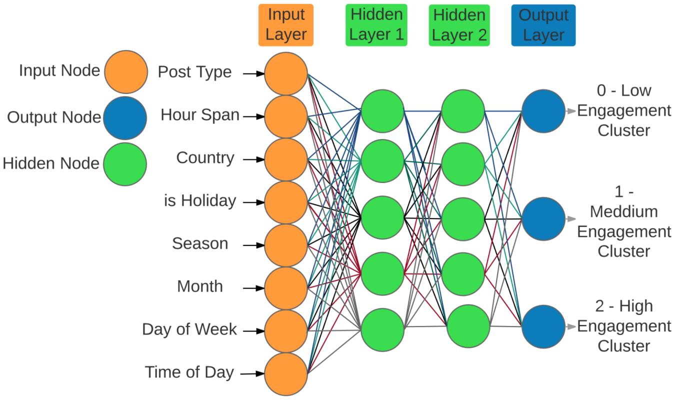
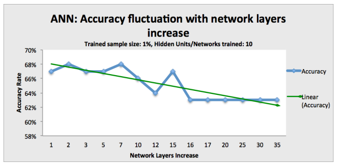


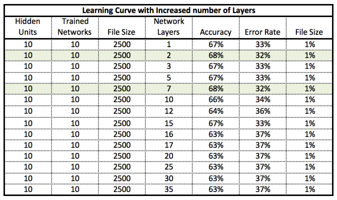
Figure 4: Neural Network with 2 layers and 5 hidden units in each layer.

2) *Deep Neural Network Structure and Learning Results:* There are numerous studies mentioned in Related work section that argue for Deep neural network (DNN) to show improved performance in comparison to ANN and therefore 3rd hypothesis was tested with the current data set. Algorithm ran 14 times, while adding few additional layers each time (from 1 to 35). All other attributes such as 10 *hidden units*, 10 *networks*

and *file size* were kept unchanged.

Figure and Table VI show linear decline when additional hidden layers are added to the network.





Learning Curve with Increased number of Layers						
Hidden Units	Trained Networks	File Size	Network Layers	Accuracy	Error Rate	File Size
10	10	2500	1	67%	33%	1%
10	10	2500	2	68%	32%	1%
10	10	2500	3	67%	33%	1%
10	10	2500	5	67%	33%	1%
10	10	2500	7	68%	32%	1%
10	10	2500	10	66%	34%	1%
10	10	2500	12	64%	36%	1%
10	10	2500	15	67%	33%	1%
10	10	2500	16	63%	37%	1%
10	10	2500	17	63%	37%	1%
10	10	2500	20	63%	37%	1%
10	10	2500	25	63%	37%	1%
10	10	2500	30	63%	37%	1%
10	10	2500	35	63%	37%	1%

Table VI: Development with hidden layer increase, File Size: 2500, 1%.

Increase in the number of layers lead to increase in the run time, as access to external GPU and large-scale distributed clusters was not available. Therefore data was trained on a smaller train sample size of 2500 lines (1% of the total data set) and tested on the rest of the data set sample. The best accuracy rate of 68% was achieved with 2 and 7 hidden layers. Additional number of layers showed linear decline. To make sure that results are not random and limited to the selected parameters, file size was increased to 14000 lines, 5% of the total data set. 2 hidden units and trained networks were selected, in order to keep the run time within realistic threshold. The best accuracy was achieved with two hidden layers and additional layers caused trend to decline.

Furthermore, file size was reduced to 0.4% and number of hidden units and trained networks fixed at 5. With gradual increase in the number of layers accuracy increased to 67% at hidden layer 5 and then fluctuated and fell to the level of the shallow network. More complex DNN structure caused over-fit with the current data set. Figure 5 shows deep neural network, that achieved accuracy of 67% with 5 hidden units in each of the 5 hidden layers, three output nodes that correspond to engagement clusters and 8 input attributes. Shallow networks perform on the same level or better than big deep neural nets when applied on public health-care data from Facebook presented here. This finding is also supported in *Do Deep Nets Really Need to be Deep?* article by [19]. Example confirms that complexity of the model does not contribute to the quality of the results in the current research, even though initially leads to better accuracy rate with the number of layers increase. This finding supports No Free Lunch Theorem by [29] who shows that in a noise-free scenario where the loss function is the misclassification rate, if one is interested in off-training-set error, then any pair of generalizers perform the same on average. Even techniques like cross validation and the use of test sets to estimate generalization error fail in as many scenarios as they succeed. Conclusively there is no learning algorithm that performs better in every case.

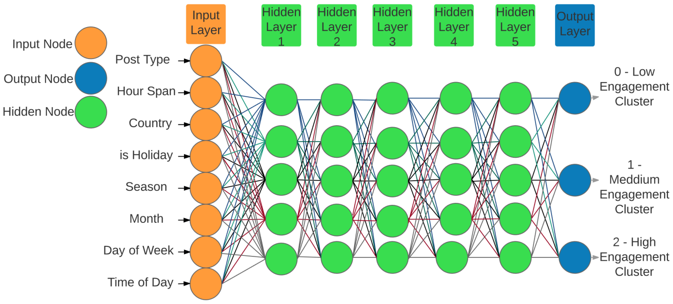


Figure 5: Deep Neural Network with 5 hidden layers

IV. Results. Post Performance prediction with ANN

Analyses from previous section suggests that there are small accuracy fluctuations with file size increase. Therefore to reduce a run time and overload on the resources best network model with five hidden units and two hidden layers will be applied on file size of 5% .

A. Prediction with ANN

Train and Test data set predictions in the next graph were necessary to see if values over-fit or not, if model memorizes the train data set rather than learning a trend. According to [28] if train set is small and number of parameters is large, model can fail to generalize and predict values that were not seen previously. Over-fitting can also occur when structure of the model does not meet the level of the noise in the data, model is too complex or due to the lack of representative samples. Usually model is easier to control than sample representation. Occam's razor principle of parsimony states: that given two models with the same generalization errors, the simpler model is preferred over the more complex one. Additional components in the more complex model stand greater chance to be fitted purely by chance [28]. Therefore, neural network with 5 hidden units in each of the 2 layers show higher accuracy results than networks with 20 or 30 hidden units. Figures 6 and 7 show classification matrix, that displays prediction results based on the total 8 attributes and neural network model with 5 hidden units in each of the 2 hidden layers and trained sample size of 5%. Matrix shows classification results both for test set and train sets and displays 'normalized' and 'not normalized' values with accuracy rate of 68%. Class 0 - low engagement cluster, is predicted with the accuracy of 90% as contains the highest amount of data points in the set. Class 1 - medium engagement cluster is predicted on unknown 'Test' data with accuracy of almost 40%. Class 2 - high engagement cluster is predicted on 'Test' data with accuracy of almost 18%. While predictions on the 'known' trained data are predicted with: around 90%, 41% and 19% accuracy rates.

Accuracy rates on 'Train' data sample set and unknown 'Test' data shows that model generalizes trend well, rather than memorizes it.

ANN classification of a single attribute suggest *Post Type*, *Hour Span* and *Time of Day* attributes to contribute the most to the classification and prediction accuracy. The rest of 5

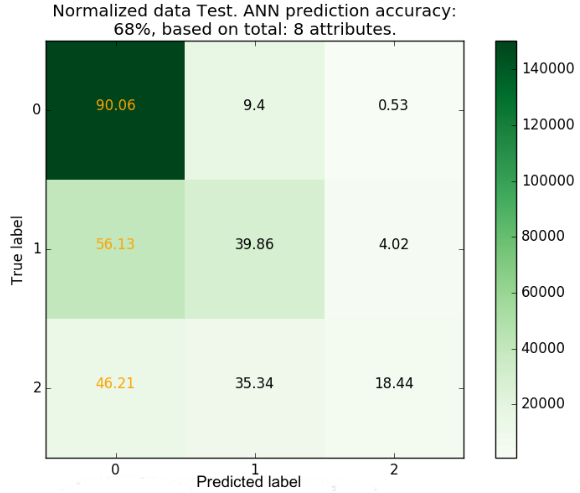


Figure 6: ANN applied on Test Data, total 8 parameters

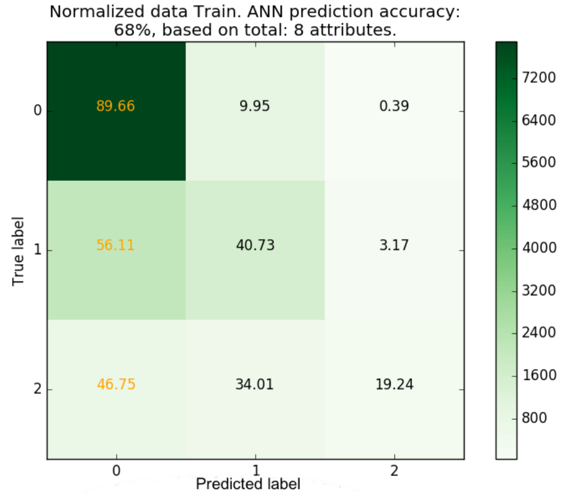


Figure 7: ANN applied on Train Data, total 8 parameters

attributes classify three ”true” labels into two classes, achieve very low accuracy rates and are too uncertain to be presented in this section.

Figure 8 shows ANN classification results for Post Type attribute (with normalized and non normalized values). ANN classifies low engagement cluster with highest accuracy of 90%, as majority of 150 thousand data points of the Test Data set are part of the *low engagement* class. Overall classification result of 68% shows that Post Type attribute has features with more distinct representation in each of the clusters than other attributes such as *isHoliday*, *Season*, *Month*, *Day of Week*, etc.

Similarly, based on *Hour Span* attribute, ANN makes prediction for data points in the *low engagement* cluster with around 93% accuracy rate, since it has the highest share representation in the *low engagement* cluster - 0. Based on the *Hour span* attribute, data in the *medium engagement* cluster is predicted with accuracy of almost 30%. Since ’true’ *high engagement* cluster has smallest share representation in the total data set, it is predicted mostly as part of the previous two clusters. Overall prediction accuracy is 65%.

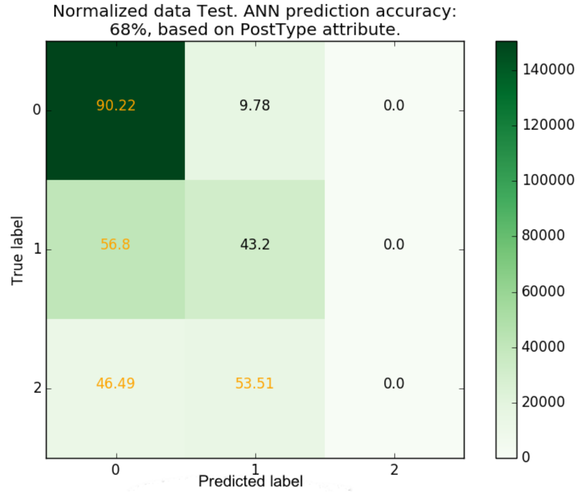


Figure 8: True Label vs Predicted Label, Post Type attribute

Time of Day attribute assigned almost 90% of ’true’ high engagement cluster into ’predicted’ medium engagement cluster. Prediction for low engagement cluster has 65% accuracy and overall prediction accuracy is also 65%.

Results from this study evidence that only three quantitative attributes can be relied on, in order to make prediction with public health care data on Facebook using ANN: *Post Type, Hour Span* and somewhat *Time of Day*. Therefore, second hypothesis is supported.

B. Engaging attributes

Majority of high engagement (44%) and medium engagement (43%) posts are posted between 10:00-16:00 and are driven by health care organization from Denmark and Norway and their corresponding work day hours. Where as greatest share of low engagement posts (37%) are posted between 16:00-20:00 and are mainly represented by US organizations.

Time laps or *Hour Span* between post creation and post update, as findings showed can contribute to higher engagement with the post. Trend showed increase in the average hour span between *Create* and *Update* date from *low engagement* to *high engagement* cluster, which might be due to post being visible on-line for longer periods of time.

Post type *Status* - short text messages achieve *Low engagement* in 65% of cases.’High’ and ’Medium’ engagement clusters contain mainly *Photo* (almost half are *High engagement* values), Video and *Link* type posts. Very often *Link Post* type contains a picture as well.

Findings suggest that visual content is of the most interest to people, who consume health care content on Facebook, followed by post types with 'link'. Analyses is based on the previous work and described in greater detail in [6].

V. Conclusion and Future Work

Optimal model used for classification, is the key to better prediction accuracy with big data sets. Even though time consuming, data preparation, pruning and right model selection will help to understand domain-specific features of the data set and contribute towards higher accuracy results. ANN prediction with quantitative data mentioned in [9] showed much lower accuracy than results achieved in this paper. Our findings show that very deep neural networks do not contribute to higher accuracy results and are quite time consuming in terms of the processing power with health care data set from Facebook. Furthermore, as part of the future work, it might be interesting to combine dynamic architecture with ANN and look into incremental clustering mentioned in [22] instead of traditional clustering approaches in order to be able to handle real-time data flow.

Current research was additionally supported with questionnaire findings that elaborated on why in spite of the higher engagement with visual content social media managers publish 'Status' type post almost 50% of the time. This is due to internal procedures and time it takes to create this type of content. Some of the more or less engaging posts contain picture and text or only text content. Choice of the key words in the title and text in the post matters, according to the expertise and suggestions of social media managers. These results will be elaborated and explored further as part of future research.

Acknowledgments

We thank the members of the Centre for Business Data Analytics (http://cssl.cbs.dk/) for their valuable feedback.

The authors were partially supported by the project *Big Social Data Analytics for Public Health: Copenhagen Health Innovation: ReVUS* funded by Region Hovedstaden and Copenhagen Commune. Any opinions, findings, interpretations, conclusions or recommendations expressed in this paper are those of its authors and do not represent the views of the funding agencies.

References

[1] H. Korda and Z. Itani, "Harnessing social media for health promotion and behavior change," *Health promotion practice*, vol. 14, no. 1, pp. 15–23, 2013. 1

[2] brandwatch. [Online]. Available: https://www.brandwatch.com/blog/47-facebook-statistics-2016/ 1

[3] Z. Tufekci, "Big questions for social media big data: Representativeness, validity and other methodological pitfalls," in *Proceedings of the Eigth International AAAI Conference on Weblogs and Social Medi*. The AAAI Press, 2014, pp. 505–514. 1

[4] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, "Predicting depression via social media." in *ICWSM*, 2013, p. 2. 1, 2

[5] M. De Choudhury, S. Counts, E. J. Horvitz, and A. Hoff, "Characterizing and predicting postpartum depression from shared facebook data," in *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. ACM, 2014, pp. 626–638. 1

[6] N. Straton, K. Hansen, R. R. Mukkamala, A. Hussain, T.-M. Grønli, H. Langberg, and R. Vatrapu, "Big social data analytics for public health," in *IEEE Healthcom´ 16*, 2016. 1, 8

[7] G. Szabo and B. A. Huberman, "Predicting the popularity of online content," *Communications of the ACM*, vol. 53, no. 8, pp. 80–88, 2010. 2

[8] M. Ghiassi, J. Skinner, and D. Zimbra, "Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network," *Expert Systems with applications*, vol. 40, no. 16, pp. 6266–6282, 2013. 2

[9] R. Sharda and D. Delen, "Predicting box-office success of motion pictures with neural networks," *Expert Systems with Applications*, vol. 30, no. 2, pp. 243–254, 2006. 2, 5, 8

[10] S. Asur and B. A. Huberman, "Predicting the future with social media," in *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on*, vol. 1. IEEE, 2010, pp. 492–499. 2, 5

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105. 2

[12] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *European Conference on Computer Vision*. Springer, 2014, pp. 818–833. 2

[13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014. 2

[14] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural computation*, vol. 18, no. 7, pp. 1527–1554, 2006. 2

[15] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015. 2

[16] D. Ciregan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 3642–3649. 2

[17] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9. 2

[18] D. C. Ciresan, U. Meier, L. M. Gambardella, and J. Schmidhuber, "Deep big simple neural nets excel on handwritten digit recognition," *CoRR*, vol. abs/1003.0358, 2010. [Online]. Available: http://arxiv.org/abs/1003.0358 2

[19] J. Ba and R. Caruana, "Do deep nets really need to be deep?" in *Advances in neural information processing systems*, 2014, pp. 2654–2662. 2, 6

[20] A. Tatar, J. Leguay, P. Antoniadis, A. Limbourg, M. D. de Amorim, and S. Fdida, "Predicting the popularity of online articles based on user comments," in *Proceedings of the International Conference on Web Intelligence, Mining and Semantics*. ACM, 2011, p. 67. 2

[21] A. Pal and S. Counts, "Identifying topical authorities in microblogs," in *Proceedings of the fourth ACM international conference on Web search and data mining*. ACM, 2011, pp. 45–54. 2

[22] H. Becker, M. Naaman, and L. Gravano, "Learning similarity metrics for event identification in social media," in *Proceedings of the third ACM international conference on Web search and data mining*. ACM, 2010, pp. 291–300. 2, 8

[23] A. Hussain and R. Vatrapu, "Social data analytics tool (sodato)," in *DESRIST-2014 Conference (in press)*, ser. Lecture Notes in Computer Science (LNCS). Springer, 2014. 3

[24] M. admin. Pearson correlation and coefficient of determination. [Online]. Available: http://mathbits.com/MathBits/TISection/Statistics2/correlation.htm 3

[25] N. J. Nagelkerke, "A note on a general definition of the coefficient of determination," *Biometrika*, vol. 78, no. 3, pp. 691–692, 1991. 3

[26] S. Wagner and D. Wagner, *Comparing clusterings: an overview*. Universität Karlsruhe, Fakultät für Informatik Karlsruhe, 2007. 3

[27] A. Rosenberg and J. Hirschberg, "V-measure: A conditional entropy-based external cluster evaluation measure." in *EMNLP-CoNLL*, vol. 7, 2007, pp. 410–420. 3, 4

[28] T. Pang-Ning, M. Steinbach, V. Kumar *et al.*, "Introduction to data mining," in *Library of congress*, vol. 74, 2006. 4, 6

[29] D. H. Wolpert, "The supervised learning no-free-lunch theorems," in *Soft computing and industry*. Springer, 2002, pp. 25–42. 6