How efficient is Twitter: Predicting 2012 U.S. presidential elections using Support Vector Machine via Twitter and comparing against Iowa Electronic Markets

Abbas Attarwala[1,2], Stanko Dimitrov[2], Amer Obeidi[2]
[1]Computer Science, Boston University, Boston, MA, 02215
[2]Department of Management Sciences, University of Waterloo, Waterloo, ON N2L6C1

***Abstract*—We test the efficient market hypothesis to see if Twitter aggregates information faster than a real-money prediction market. We use Support Vector Machines (SVMs), a supervised learning algorithm, to predict the outcome of the 2012 U.S. presidential elections via Twitter data. We then compare the prediction from SVM against the Iowa Electronic Markets (IEM). A total of 40 million unique tweets were collected and analyzed between September 29^{th} 2012 and November 6^{th} 2012. We observe: 1) the IEM is efficient on all the above days as per the semi-strong efficient market hypothesis definition [1]. SVM does not out predict the IEM. 2) The SVM prediction results are positively correlated with the IEM and predicts Obama winning the election, implying that Twitter can be considered as a valid source in predicting US presidential election outcomes. Using the Granger causality test, no causal relationship was inferred between the two-time series. 3) The candidate frequency count distribution independent of any sentiment analysis on all days is also positively correlated with IEM and SVM. Using Granger causality test, we determined that IEM statistically causes the candidate frequency count distribution in Twitter at the 1% level.**

***Keywords*—Support Vector Machine; Sentimental Analysis; Prediction Market; Efficient Market Hypothesis**

I. Introduction

In a truly efficient election market, the market price is the best predictor of an event [2]. No other information can be used to improve the market generated forecast. The price is a sufficient statistic for all private and public information held by traders. This paper attempts to address four related research questions:

1) Is the IEM efficient as per the semi-strong efficient market hypothesis proposed by Fama [1]?
    a) If the IEM is efficient, then an SVM will not out predict it.
    b) However if an SVM predicts the IEM does this imply:
        i) Traders in the IEM are not aware of underlying pattern(s)/trend(s) on Twitter?
        ii) Do arbitrage opportunities exist in the IEM?

    Traditional media such as radio, television, newspapers and books are static in space and time [3]. This is in complete contrast with the dynamism of social media that evolve continuously across space and time. Social media has transformed these traditional channels in numerous ways. For example, Twitter, Flickr and online collaboration on Google Maps have changed the way in which knowledge and news such as the Hudson river accident in 2009 and finding survivors during the 2004 Indonesia earthquake disseminate among people. Information is dynamic and a function of spatial and temporal properties, capturing the mood and sentiment of people. Researchers [4], [5] have pointed out that celebrities and politicians use social platforms to enhance their star and political power. Celebrities promote their upcoming movies on social media so that it may increase its box office revenue. Politicians promote their political agenda and in turn tend to increase their support base and their prospects of winning elections. This paper uses Twitter as its data source, and addresses the following research questions in addition to the one previously mentioned:
2) Is Twitter efficient as per the Efficient Market Hypothesis in predicting the 2012 U.S. presidential elections?
3) Is the prediction of U.S. 2012 presidential election using SVMs comparable to IEM?
4) Is there any causal relationship between the IEM and the best found SVM predictions? Likewise, is there any causal relationship between the IEM prediction and the frequency count distribution of tweets mentioning either presidential candidate? We use the Toda Yamamoto [6] version of GCT in this paper.

In the remainder of this paper, we first discuss related work. Next, we describe the model and the problem setup. We then discuss the analysis and results. Finally, we conclude and present future work.

II. Related Work

Gaurav et al. [7] use Twitter data to predict the election outcome in Latin American countries. Their prediction is based on the number of times the name of the candidate is mentioned in tweets prior to elections. The authors use methods to augment the counts by counting not only the presence of candidate's official names but also their aliases and commonly appearing names.

Researchers [8] use the context of the German federal election to investigate whether Twitter can predict German elections? The authors use text analysis software [9]; and conducted content analysis of over $100,000$ messages containing a reference to either a political party or a politician. Tumasjan et al. [8] find that the mere number of messages mentioning a party reflects the election result. We find similar results with the U.S. presidential elections. We infer that the number of tweets mentioning either of the candidate independent of any sentimental analysis is a very good reflection of the U.S. presidential outcome.

Mishne and Glance's main finding in their research [10] is that positive sentiment is a better predictor for movie success when applied to limited context around reference to the movie in online posts, posted prior to its release. We in our paper use a similar technique of finding the ratio of positive sentiment/negative sentiment for each candidate on all positive sentiment/negative sentiment tweets and use this metric in predicting the winning candidate. Section 5 of our paper discusses this in greater detail.

The paper [11] focuses on detecting events popularity through sentiment analysis of tweets published by the financial community on the Twitter universe. The researchers filter out all the noisy tweets in order to analyze only the tweets that influence the financial market. Researchers [12] try to predict stock market indicators such as Dow Jones, NASDAQ and S&P 500 by analyzing Twitter posts. The authors inferred 'hope' and 'fear' sentiment using Twitter data. They found that emotional tweet percentage is significantly negatively correlated with Dow Jones, NASDAQ and S&P 500, but displayed significant positive correlation to VIX, a volatility measure.

Johan and Mao [13] address whether societies experience mood states that affect their collective decision making? Is the public mood correlated or even predictive of economic indicators? The authors study whether measurements of collective mood states derived from large-scale Twitter feeds are correlated to the value of the Dow Jones Industrial Average (DJIA) over time. Finally, the authors cross-validate the resulting mood time series by comparing their ability to detect the public's response to the presidential election and Thanksgiving day in 2008. Sitaram and Huberman [5] demonstrate how social media content can be used to predict real-world outcomes. They use Twitter posts, to forecast box-office revenues for movies. Further, they conclude that using a simple model built from the rate at which tweets are created about particular topics can outperform market-based predictors. The preceding financial market-Twitter studies are quite interesting, but the underlying financial time-series may have other drivers besides those extracted from Twitter. As such, it is unclear if the same results naturally carry over to a prediction-market time series.

Jahanbakhsh [14] examines the predictive power of Twitter regarding the U.S. presidential election of 2012. In their research, they analyzed 32 million tweets regarding the U.S. presidential election by employing counting techniques and Naive Bayes model. They devised an advanced classifier for sentiment analysis to increase the accuracy of Twitter content analysis and validated their results with comparison to traditional opinion polls. However, the authors do not compare the efficiency of their predictions to election markets like we do in our paper.

Researchers [15] have performed large-scale Twitter sentiment analysis to examine the orthodontic patient experience having braces compared with Invisalign. These researchers collected tweets containing the words "braces" or "Invisalign" for a period of 5 months. In our research as part of data gathering process, our tweets contained either the words "Romney" or "Obama".

Chung and Mustafaraj [16] have mentioned that it may be necessary in future research to perform preprocessing techniques such as Part of Speech (POS) tagging in improving the accuracy of sentiment analysis on Twitter data. We incorporate these techniques as mentioned in section Preprocessing and Part of Speech Tagging (POS) of our paper.

Most researchers as cited above, have resorted to simple counting techniques using Naive Bayes for prediction. It has been pointed out [17], [18] that Naive Bayes has limitations when dealing with linguistic and text based classifications. To our knowledge, no researchers used SVMs to predict the 2012 U.S. presidential elections with a tweet corpus of 27 million and compared the predictions to the Iowa presidential market results. We use the radial bias function kernel in our SVM. Our research exploits the hidden semantic structure and augments linguistics dimensions by using part of speech tagging (POS) in the feature space using Stanford NLP library [19].

III. Problem Setup and Model

A. Problem Setup

In order to address our research questions, we first collected Twitter data, commonly referred to as tweets, from September 29th 2012 to November 6th 2012. Let $T = \{t_i, .., t_e\}$ denote the set of days where t_i refers to starting day of September 29th 2012 and t_e refers to end date of November 6th 2012. Approximately 40 million tweets were collected using Twitter's restful API. From these 40 million tweets, only tweets that mentioned exactly one of the two candidates (either Obama or Romney) were selected. When both candidates are present in the tweet, it is more challenging to compute the right sentiment of the tweet. This pruning resulted in a reduced number of tweets of approx 27 million. Let P denote set of tweets:

$$P = \left\{tw_{t_i,1}, ..tw_{t_j,k}, ..tw_{t_e,27\ million}\right\}. \tag{1}$$

Where $tw_{l,k}$ is tweet on day l. $tw_{l,k}$ is the raw unstructured text without any preprocessing and collected directly from Twitter. It follows that P is the list of tweets temporally arranged. Where $t_i \leq t_j \leq t_e$ and $1 \leq k \leq 27\ million$.

1) Training Data: We use SVM in a supervised learning setting. In order to train a supervised learning algorithm, a set of labeled data is required. For this paper, we outsourced labelling of 3000 random tweets, selected between September 25th 2012 and September 28th 2012.

$$TrainingData = \{(tw_{t,i}, s_i) \mid s_i \in \{+1, 0, -1\}\}_{i=1}^{3000}. \tag{2}$$

For a given tweet $tw_{t,i}$ in the above 4 days, a sentiment, s_i is attached of either $+1$=positive, 0=neutral or -1=negative. As an example, the following labels were ssigned for the following two tweets:

> (1)—"@steverocknroll thats fine,i just told you my opinion based on my observations, Obama cannot please everyone"
> obama=+1,romney=na was attached.
> (2)—"....wait, Romney actually has a chance to become President...I'm moving to Canada."
> obama=na,romney=-1

2) Preprocessing and Part of Speech Tagging (POS): All tweets in the set P and $TrainingData$ were first preprocessed and normalized. For each tweet the following were performed as part of preprocessing [20]:

1) All tokens and words were turned into lower case.
2) Characters such as @ and # were removed.
3) Sentence break [21] were implemented and performed.
4) HTML character codes and clitics were cleaned up and removed.
5) URLs were removed.

In corpus linguistics, POS [20] tagging is the process of marking a token in a corpus with its corresponding part of speech, based on both its definition, as well as its context. Using the Stanford NLP library, POS was then performed on each token in the tweet. Each token is appended with a / followed by the POS. The two sample tweets above, after preprocessing and POS tagging, resulted in the following:

> (1)—"steverocknroll/NN thats/NNS fine/NN ,/NN i/NN just/NN told/NN you/NN my/NN opinion/NN based/VBN on/NN my/NN observations/NNS ,/NN Obama/NN cannot/NN please/NN everyone/NN"
> (2)—".../CD ./CD wait/NN ,/NN Romney/NN actually/RB has/NNS a/NN chance/NN to/NN become/NN President/NN ../CD ./CD I/NN 'm/NN moving/VBG to/NN Canada/NN ./CD"

3) ARFF file for WEKA: Each normalized and preprocessed tweet $\mathbf{x_j}$ is a vector $\in \mathbb{R}^{25}$. This feature space consists of:

- Unique POS count. There are total of 20 unique POS.
- Average sentence length.
- Number of top election keywords [22].
- Number of positive sentiment keywords and number of negative sentiment keywords [23].
- Number of positive tokens per sentence and number of negative tokens per sentence.

Each tweet vector was formatted as per the ARFF (Attribute-Relation File Format) [24] file specification. An ARFF file is an ASCII text file that describes a list of data vectors sharing a set of attributes.

4) Creating Support Vector Machine Model from Training Data: A multivariate classifier across three sentiments $\{+1, 0, -1\}$ using SVM, is simply three hyperplanes, each acting as a binary classifier on the tuple sentiment of $(+1, 0)$, $(+1, -1)$ and $(0, -1)$.

A general two class (+1,-1) classification problem using SVM is of the linear form, for transformation function $\phi(\cdot)$:

$$y_{(+1,-1)}(\mathbf{x}) = \mathbf{w^T}\phi(\mathbf{x}) + b, \qquad (3)$$
$$\phi(\mathbf{x}) \in \mathbb{R}^D, D > 25, \mathbf{x} \in \mathbb{R}^{25}.$$

All $\mathbf{x_j}$ in class $s_j = +1$ have $y(\mathbf{x_j}) = +1$, the analogous holds for $s_j = -1$. In order to determine the values $\mathbf{w}$ and b, the following optimization problem is solved:

$$\arg\min_{\mathbf{w},b} \quad \frac{||\mathbf{w}||^2}{2} + C\sum_{i=1}^{n}\zeta_i \qquad (4)$$
$$\text{s.t.} \quad s_n(\mathbf{w^T}\phi(\mathbf{x_n}) + b) \geq 1 - \zeta_i \qquad (5)$$
$$\zeta_i \geq 0 \qquad \forall n \in 1, \ldots, N, .$$

Where N is the number of data points to be classified. It can be shown that solving system (4)–(5) is equivalent to solving a quadratic optimization problem for a select few $\phi(\cdot)$ functions, where $K(\mathbf{x_i}, \mathbf{x_j}) = \phi(\mathbf{x_i})^T\phi(\mathbf{x_j})$ is referred to as the *kernel function*. We consider the radial bias kernel function in this manuscript, where $K(\mathbf{x_i}, \mathbf{x_j}) = e^{\frac{-||\mathbf{x_i}-\mathbf{x_j}||^2}{2\sigma^2}}$ or $K(\mathbf{x_i}, \mathbf{x_j}) = e^{-\gamma||\mathbf{x_i}-\mathbf{x_j}||^2}, \gamma > 0$. ζ_i are the slack variables that allow an example to be in the margin $0 \leq \zeta_i \leq 1$ or be misclassified $\zeta > 1$ [25]. $\sum_i \zeta_i$ is a bound on the number of misclassified examples. $C\sum_i \zeta_i$ is used to penalize misclassification and margin errors in the objective function (4).

5) Classifying new tweets: In order to classify new tweets $\mathbf{x} \in P$, using the trained model we determine the sign of $y(\mathbf{x})$.

WEKA software calculates the probabilities across the three sentiments of $\{+1, 0, -1\}$ using Platt scaling [26]. Platt scaling is an algorithm used to derive probability estimates by fitting a logistic regression model to the SVM classifier results:

$$\Pr(s = +1|\mathbf{x}, y_{(+1,-1)}(\mathbf{x})) = \frac{1}{1 + e^{Ay_{(+1,-1)}(\mathbf{x})+B}}. \qquad (6)$$

A logistic transformation of the classifier scores $y_{(+1,-1)}(\mathbf{x})$, where A and B are scalar parameters learned by the algorithm. Note that classifications are now determined by

$$s = +1 \; if \Pr(s = +1|\mathbf{x}) > \frac{1}{2}; \; and \; if \; B \neq 0.$$

To find the probability of negative sentiment on the binary classification problem of $(+1, -1)$ is:

$$\Pr(s = -1|\mathbf{x}, y_{(+1,-1)}(\mathbf{x})) = 1 - \Pr(s = +1|\mathbf{x}, y_{(+1,-1)}(\mathbf{x})).$$

Note: SVM built for this project, is a multivariate classifier, we have three such hyperplanes: $y_{(+1,-1)}(\mathbf{x}), y_{(+1,0)}(\mathbf{x}),$ and $y_{(0,-1)}(\mathbf{x})$.

For a given new tweet $\mathbf{x}$, the sentiment is calculated as:

$$\arg\max_s \Pr(s \mid \mathbf{x}),$$

where s is a random variable and can take on values from the set $\{+1, 0, -1\}$. $+1$ refers to positive sentiment, 0 refers to neutral sentiment and -1 refers to negative sentiment.

IV. Analysis

A. Hyper-parameter tuning

Hyper-parameters are outside of standard model training and are typically predefined. There are two hyper-parameters in the SVM model. These are γ and C. It is not known beforehand which C and γ are best for a given problem; consequently some kind of model selection must be done [27]. We use the "grid-search" on C and γ using cross validation. As mentioned [27], various pairs of (C, γ) values are tried and the one with the best cross-validation accuracy is picked. We used the following ranges on C and γ for identifying the parameter values: $C : 2^{-5}, 2^{-3}, \ldots, 2^{15}$ and $\gamma : 2^{-15}, 2^{-13}, \ldots, 2^{3}$. Using a step cost of 2 we exhaustively try every pair, i.e., a total of 90. For every such pair of γ and C we execute 10 fold cross validation and then select the best hyper-parameters value $(C = 3, \gamma = .0078125)$ that yield the highest cross validation rate, i.e., 51%. The optimal neighborhood is then further refined by re-executing "grid-search", but now using a smaller step cost of .5 and limiting the lower and upper bounds of $\gamma : 2^{-9}, \ldots, 2^{-6}$ and $C : 2^{1}, \ldots, 2^{5}$ in the optimal neighborhood. We use $(C = 3, \gamma = .005)$ that yields the cross validation rate of approximately 53%. We use exhaustive search, however some [28] have used optimization on hyper-parameter search using random search and greedy sequential methods when training neural networks and deep belief networks. The following table lists the detailed accuracy of the trained model using standard metrics. The last row in the table is the weighted average across the three classes.

True Positive Rate	False Positive Rate	Precision	Recall	F-Measure	ROC Area	Class
.027	.002	.8	.027	.052	.527	positive
.6	.278	.594	.6	.597	.668	negative
.706	.505	.478	.706	.57	.601	neutral
.528	.313	.589	.528	.478	.613	

TABLE I. Results across the three classes

B. Results and Discussions

1) Temporal distribution of all tweets: Figure 1, shows the temporal distribution of all the tweets collected for this project. The y axis represents the number of tweets collected and the x axis represents the date. The figure shows three local maxima (October 3rd, October 16th and October 22nd) and a global maximum on election date (6th November 2012). The three local maxima correspond to presidential debates.

2) Count distribution of Obama or Romney mention in tweet: Figure 2, shows the temporal distribution of tweet mentioning either Obama or Romney but not both. This analysis was carried out on all tweets. It is interesting to note that between 3rd October and 20th October, Obama led Romney slightly: max Obama of 53% with a max Romney of 51% and a min Obama of 48% with a min Romney of 44%. However, after the conclusion of the presidential debates the lead widens in favor of Obama. Around 25th October and leading upto the election date of 6th November, the two time series have diverged with Obama having a max of 77% and Romney having a max of 48%. We like to remind the reader again, that no sentiment analysis is carried out yet. This is simply the temporal distribution of tweets either mentioning Obama or Romney independent of whether the tweet is classified as positive, negative or neutral.

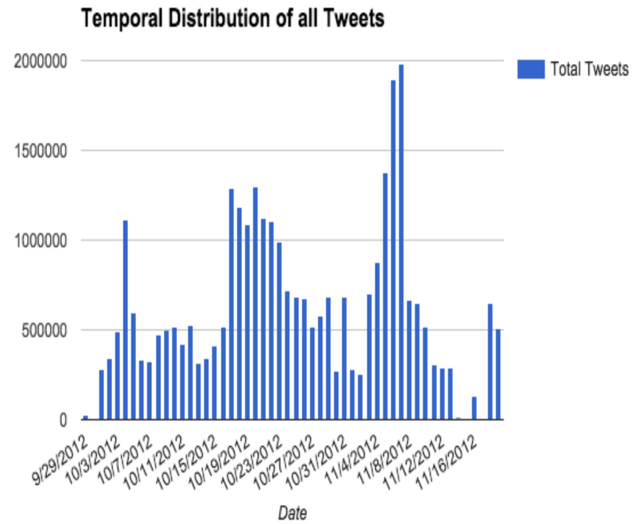

Fig. 1. Temporal distribution of all tweets collected.

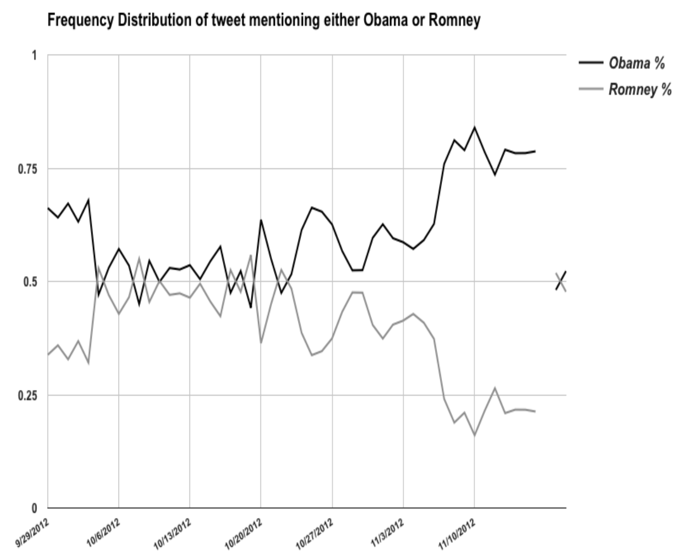

Fig. 2. Frequency distribution of Obama or Romney mention in a tweet.

3) Iowa Electronic Markets Results: IEM [29] consists of two kinds of securities/markets:

1) Winner Takes All Market (WTA): The financial contracts traded in this market are DEM12_WTA and REP12_WTA with the following pay off: The holder of DEM12_WTA (REP12_WTA) receives \$1 if the Democratic Party (Republican Party) nominee receives the majority of popular votes cast for the two major parties in the 2012 U.S Presidential elections, \$0 otherwise.
2) Vote Share (VS): Payoffs in the 2012 Presidential Vote-Share Market will be determined by the % of the popular vote received by the official Democratic and Republican nominees in the 2012 U.S. Presidential election. Specifically, payoffs are determined by the % of the total two-party popular vote received by each

of the two parties in 2012 U.S Presidential election.

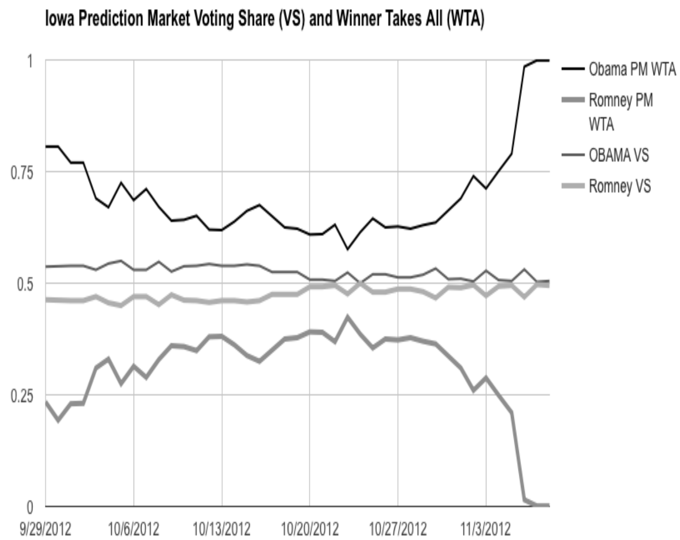

Fig. 3. IEM of popular vote share (VS) and Winner Takes All (WTA)

Results for IEM using WTA and VS are shown in Figure 3.

4) Election Market vs Obama/Romney mention in Tweet: The correlation coefficients are summarized in Table II.

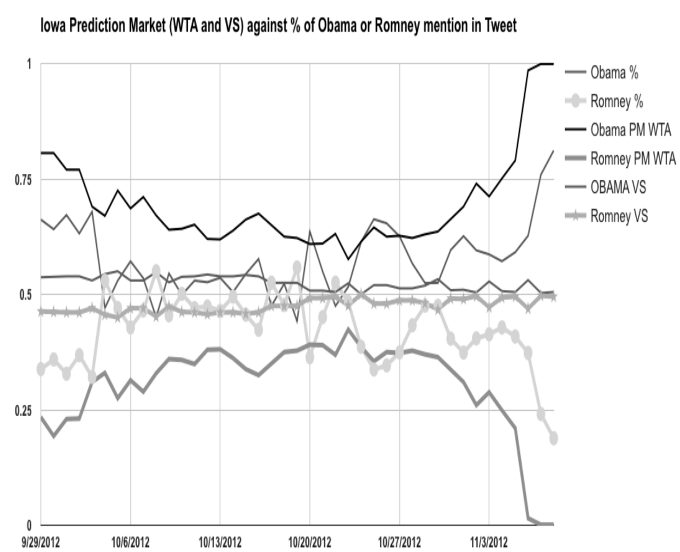

Fig. 4. IEM with Frequency distribution of Obama or Romney mention in a tweet

	ObamaWTA	ObamaTwitterCount	RomneyTwitterCount	RomneyWTA
ObamaWTA	1.0000	.7433	-.6829	-.9994
ObamaTwitterCount	.7433	1.0000	-.9653	-.7413
RomneyTwitterCount	-.6829	-.9653	1.0000	.6807
RomneyWTA	-.9994	-.7413	.6807	1.0000

TABLE II. MATRIX OF CORRELATION COEFFICIENTS OF WTA WITH CANDIDATE MENTION IN TWEET

A time series X is said to Granger-cause Y if it can be shown, through F-tests on lagged values of X (and with lagged values of Y also included), that those X values provide statistically significant information about future values of Y [30]. By using first differences, both the time series (IEM WTA and Obama/Romney mention in tweet) were converted into stationary processes. GCT is now performed after Augmented Dicky-Fuller (ADF) [31] to determine whether:

- Iowa WTA can be used to forecast the Obama/Romney mention in tweet time series?
- Obama/Romney mention in tweet time series can be used to forecast the Iowa WTA time series.?

α is set to 1% and the maximum number of lags is set to 5 days. The null hypothesis H_0 is Y does **not** Granger cause X. If the H_0 is rejected, then the optimal lag length for X and Y based on Bayesian Information Criterion is determined. Results from GCT are as follows: Using the

	ObamaWTA (Y)	%Obama in Twitter (Y)
ObamaWTA (X)		FValue=5.3181 CValue=3.8575 i.e. Reject Null Hypothesis
%Obama in Twitter (X)	FValue=4.9544,CValue=6.5030 i.e. Do not reject Null Hypothesis	

TABLE III. GCT FOR OBAMA WTA AND OBAMA % MENTION IN TWITTER

	RomneyWTA (Y)	%Romney in Twitter (Y)
RomneyWTA (X)		FValue=5.3181 CValue=3.8575 i.e. Reject Null Hypothesis
%Romney in Twitter (X)	FValue=4.9544 CValue=6.5030 i.e. Do not reject Null Hypothesis	

TABLE IV. GCT FOR ROMNEY WTA AND ROMNEY % MENTION IN TWITTER

GCT (see Table III and Table IV), we can statistically conclude at $\alpha = 1\%$ that the Iowa WTA electronic market causes the Obama/Romney mention-in-tweet time series. The optimal lag length is determined 5 days indicating that it takes up to 5 days after the contracts are traded on Iowa WTA before the information is reflected in tweets.

5) Support Vector Machine Predictions: Figure 5 shows the predicted probability of each candidate winning the election as determined by our trained SVM model. In order to calculate the probabilities of either candidate winning, we first calculated the following counts across all tweets in P for days starting from September 29th 2012 to November 6th 2012.

$$Count(Candidate = Obama, Sentiment = +Sentiment, day_j) \quad (7)$$

$$Count(Candidate = Obama, Sentiment = -Sentiment, day_j) \quad (8)$$

$$Count(Candidate = Romney, Sentiment = +Sentiment, day_j) \quad (9)$$

$$Count(Candidate = Romney, Sentiment = -Sentiment, day_j) \quad (10)$$

For each day_j, we define the probabilities of winning using equation (7)–(10) as follows:

$$\Pr(Candidate = Obama, day = day_j) = \frac{\frac{(7)}{(8)}}{\frac{(7)}{(8)} + \frac{(9)}{(10)}}, \quad (11)$$

$$\Pr(Candidate = Romney, day = day_j) = \frac{\frac{(9)}{(10)}}{\frac{(9)}{(10)} + \frac{(7)}{(8)}}. \quad (12)$$

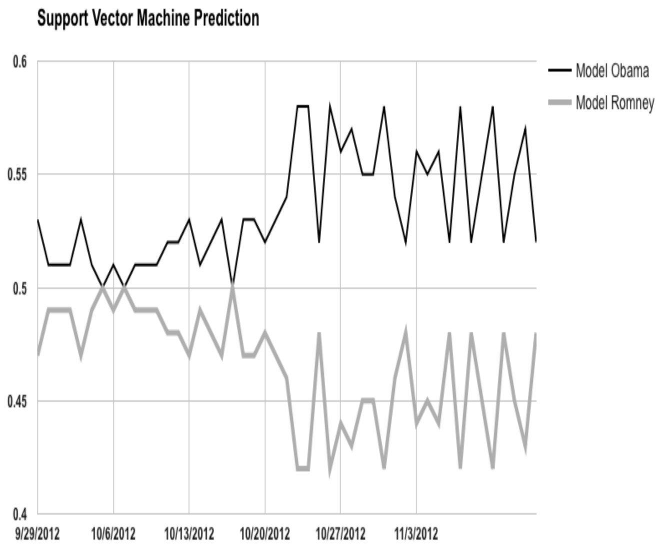

Fig. 5. Prediction using Support Vector Machine

The coefficient correlation matrix with IEM WTA market is summarized in Table V.

	ObamaSVM	ObamaWTA	RomneySVM	RomneyWTA
ObamaSVM	1.0000	.4772	-1.0000	-.4854
ObamaWTA	.47772	1.0000	-.4772	-.9992
RomneySVM	-1.0000	-.4772	1.0000	.4854
RomneyWTA	-.4854	-.9992	.4854	1.0000

TABLE V. MATRIX OF CORRELATION COEFFICIENTS OF WTA WITH SVM PREDICTIONS

ADF test on time series data is executed twice: a) prior to first differences and then again b) after first differences to assess whether the series have a unit root and for stationary test. GCT is now performed for determining whether:

- WTA can be used to forecast the SVM predictions?
- SVM predictions can be used to forecast the WTA time series?

We set α to 1%, and the maximum number of lags is set to 5 days for GCT. The null hypothesis H_0 is Y does **not** Granger cause X. Results from GCT are as follows, in tables VI and VII:

	ObamaWTA (Y)	Obama SVM (Y)
ObamaWTA (X)		FValue=3.9011 CValue=7.2339 i.e. Do not reject Null Hypothesis
Obama SVM (X)	FValue=5.0046 CValue=6.5030 i.e. Do not reject Null Hypothesis	

TABLE VI. GCT FOR OBAMA WTA AND OBAMA SVM

	RomneyWTA (Y)	Romney SVM (Y)
RomneyWTA (X)		FValue=3.9011 CValue=7.2339 i.e. Do not reject Null Hypothesis
Romney SVM (X)	FValue=4.1583 CValue=4.2726 i.e. Do not reject Null Hypothesis	

TABLE VII. GCT FOR ROMNEY WTA AND ROMNEY SVM

Using the GCT, we can statistically conclude at $\alpha = 1\%$ that no causal relationship may be established between the two time series (SVM predictions and IEM WTA market). The conclusion holds even when α is set to 5% and 10%.

V. CONCLUSION

Using correlation coefficients and GCT, we conclude that the IEM WTA causes the frequency count of tweet mentioning either Obama or Romney. However, no such causal relationship were inferred from the time series data of IEM WTA and SVM predictions. Correlation coefficients suggest that SVM predictions are positively related to the IEM. We have not accounted for demographics, location, age, social and racial groups in Twitter data. For instance, younger audience and people from urban areas will be overrepresented in the Twitter data [32]. There is also a strong bias of Twitter audience leaning towards Democratic candidates [33]. For future work in this area, it is essential to address these potential biases.

We do not claim in this paper that Iowa Electronic Markets out predicts all supervised learning algorithms. There may be other supervised or unsupervised learning algorithms that may out predict IEM. The purpose of this paper is to understand better the research questions outlined in the introduction in the context of SVMs. SVMs does not out predict the IEM on any of the days between September 29th, 2012 and November 6th, 2012. However, the SVM is positively related to Iowa WTA prediction market and results in Obama winning the election 58% on election day compared to 95% with Iowa WTA prediction market and 53% with Iowa vote share market.

ACKNOWLEDGMENTS

We like to acknowledge and thank Kazem Jahanbakhsh at the University of British Columbia, for sharing his Twitter corpus on U.S. 2012 presidential elections with us. We also acknowledge the financial support we received to conduct this work from the Province of Ontario, the University of Waterloo, and the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

[1] E. F. Fama, "Efficient capital markets: A review of theory and empirical work," *The journal of Finance*, vol. 25, no. 2, pp. 383–417, 1970.

[2] J. Wolfers and E. Zitzewitz, "Prediction markets," *The Journal of Economic Perspectives*, vol. 18, no. 2, pp. 107–126, 2004.

[3] S. MacBride, *Many voices, one world: Towards a new, more just, and more efficient world information and communication order*. Rowman & Littlefield, 1980.

[4] W. G. Mangold and D. J. Faulds, "Social media: The new hybrid element of the promotion mix," *Business horizons*, vol. 52, no. 4, pp. 357–365, 2009.

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

[6] D. Theodoridis, "The granger causal relationship between energy consumption and economic growth for eight european countries," 2013. Student Paper.

[7] M. Gaurav, A. Srivastava, A. Kumar, and S. Miller, "Leveraging candidate popularity on twitter to predict election outcome," in *Proceedings of the 7th Workshop on Social Network Mining and Analysis*, p. 7, ACM, 2013.

[8] A. Tumasjan, T. Sprenger, P. Sandner, and I. Welpe, "Predicting elections with twitter: What 140 characters reveal about political sentiment," in *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, pp. 178–185, 2010.

[9] J. W. Pennebaker, M. E. Francis, and R. J. Booth, "Linguistic inquiry and word count: Liwc 2001," *Mahway: Lawrence Erlbaum Associates*, vol. 71, p. 2001, 2001.

[10] G. Mishne and N. S. Glance, “Predicting movie sales from blogger sentiment.,” in *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs*, pp. 155–158, 2006.

[11] M. Daniel, R. F. Neves, and N. Horta, “Company event popularity for financial markets using twitter and sentiment analysis,” *Expert Systems with Applications*, vol. 71, pp. 111–124, 2017.

[12] X. Zhang, H. Fuehres, and P. A. Gloor, “Predicting stock market indicators through twitter ”i hope it is not as bad as i fear”,” *Procedia-Social and Behavioral Sciences*, vol. 26, pp. 55–62, 2011.

[13] J. Bollen, H. Mao, and X. Zeng, “Twitter mood predicts the stock market,” *Journal of Computational Science*, vol. 2, no. 1, pp. 1–8, 2011.

[14] K. Jahanbakhsh and Y. Moon, “The predictive power of social media: On the predictability of U.S. presidential elections using twitter,” *CoRR*, vol. abs/1407.0622, 2014.

[15] D. Noll, B. Mahon, B. Shroff, C. Carrico, and S. J. Lindauer, “Twitter analysis of the orthodontic patient experience with braces vs invisalign,” *The Angle Orthodontist*, 2017.

[16] J. E. Chung and E. Mustafaraj, “Can collective sentiment expressed on twitter predict political elections?,” in *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence*, AAAI Press, 2011.

[17] G. Qiang, “An effective algorithm for improving the performance of naive bayes for text classification,” in *2010 Second International Conference on Computer Research and Development*, pp. 699–701, 2010.

[18] I. Rish, “An empirical study of the naive bayes classifier,” in *IJCAI 2001 workshop on empirical methods in artificial intelligence*, vol. 3, pp. 41–46, IBM New York, 2001.

[19] C. D. Manning, “The Stanford CoreNLP natural language processing toolkit,” in *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 55–60, 2014.

[20] A. Voutilainen, “Part-of-speech tagging,” *The Oxford handbook of computational linguistics*, pp. 219–232, 2003.

[21] Y. Gotoh and S. Renals, “Sentence boundary detection in broadcast speech transcripts,” 2000.

[22] M. P. Marcus, M. A. Marcinkiewicz, and B. Santorini, “Building a large annotated corpus of english: The penn treebank,” *Computational linguistics*, vol. 19, no. 2, pp. 313–330, 1993.

[23] A. Esuli and F. Sebastiani, “Sentiwordnet: A publicly available lexical resource for opinion mining,” in *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC-2006)*, (Genoa, Italy), European Language Resources Association (ELRA), May 2006. ACL Anthology Identifier: L06-1225.

[24] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, “The weka data mining software: an update,” *ACM SIGKDD explorations newsletter*, vol. 11, no. 1, pp. 10–18, 2009.

[25] A. Ben-Hur and J. Weston, “A user’s guide to support vector machines,” *Data mining techniques for the life sciences*, pp. 223–239, 2010.

[26] J. Platt, “Fast training of support vector machines using sequential minimal optimization,” in *Advances in Kernel Methods - Support Vector Learning* (B. Schoelkopf, C. Burges, and A. Smola, eds.), MIT Press, 1998.

[27] C.-W. Hsu, C.-C. Chang, C.-J. Lin, *et al.*, “A practical guide to support vector classification,” 2003.

[28] J. S. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, “Algorithms for hyper-parameter optimization,” in *Advances in Neural Information Processing Systems 24* (J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, and K. Q. Weinberger, eds.), pp. 2546–2554, Curran Associates, Inc., 2011.

[29] J. E. Berg, F. D. Nelson, and T. A. Rietz, “Prediction market accuracy in the long run,” *International Journal of Forecasting*, vol. 24, no. 2, pp. 285–300, 2008.

[30] C. W. J. Granger, “Investigating causal relations by econometric models and cross-spectral methods,” *Econometrica*, vol. 37, no. 3, pp. 424–438, 1969.

[31] D. A. Dickey and W. A. Fuller, “Distribution of the Estimators for Autoregressive Time Series with a Unit Root,” *Journal of the American Statistical Association*, vol. 74, pp. 427–431, June 1979.

[32] M. Duggan and A. Smith, “Demographics of key social networking platforms,” 2013.

[33] D. Gayo-Avello, ““i wanted to predict elections with twitter and all i got was this lousy paper”–a balanced survey on election prediction using twitter data,” *arXiv preprint arXiv:1204.6441*, 2012.