Estimating confidence of Agricultural Producers by mining social media. Preliminary Results.

Xavier Ignacio González[1], Emilio Picasso[1]

[1] University of Buenos Aires - School of Engineering
Av. Las Heras 2214, CABA, Rep. Argentina
xgonzalez@fi.uba.ar

Abstract

This document discusses preliminary results about predicting the confidence of Agricultural Producers in Argentina by mining social media. The Confidence Index is measured every four months by means of a survey. We have developed a confidence sentiment measurement, based on social media, that evolves close to the Confidence Index. Then, the former would serve as a leading indicator of changes in the latter.

1 Introduction

Agriculture is one of the largest industries in argentine economy. Since agriculture is predominantly a private business, its development and growth depends on individual strategic decisions made by farmers. These decisions are mainly determined by the confidence producers have in relation to the economic, sectorial, and financial context.

To reflect cyclical movements of economic activity in the short term and evaluate the disposition of the sector to growth, an organization called CREA (*Consorcio Regional de Experimentacion Agícola* - Regional Consortium of Agricultural Experimentation) has developed the Confidence Index of the Agricultural Entrepreneur (ICEA - *Índice de Confianza del Empresario Agropecuario*). This indicator is a quantitative measure of agricultural entrepreneurs' perceptions about the business context, operationalized by means of an online survey implemented every four months. This index has allowed to track the evolution of the sector during five years since its creation. However, faster response would be desirable for policy evaluation. The information is there in the social media, however lacking the structure and conclusiveness of a quantitative indicator.

We develop an alternative way to measure farmers' confidence, by mining opinions in Twitter. A few studies explored the possibility of understanding public opinion through mining social media [1-7], however to the best of our knowledge, there is no previous research on social media and the confidence of agricultural producers.

2 Methodology

In summary, this study carried out the following steps: collecting tweets, filtering relevant tweets, obtaining sentiment aggregated scores, and comparing aggregated scores to the Confidence Index of Agricultural Entrepreneur (ICEA) obtained by the corresponding survey.

2.1 Collecting tweets

The time frame to select the tweets to be analyzed was set from July 2014 to July 2016. Within this period, there were significant changes in the index level due to political changes occurred in Argentina. The evolution of the index is depicted in Figure 1.

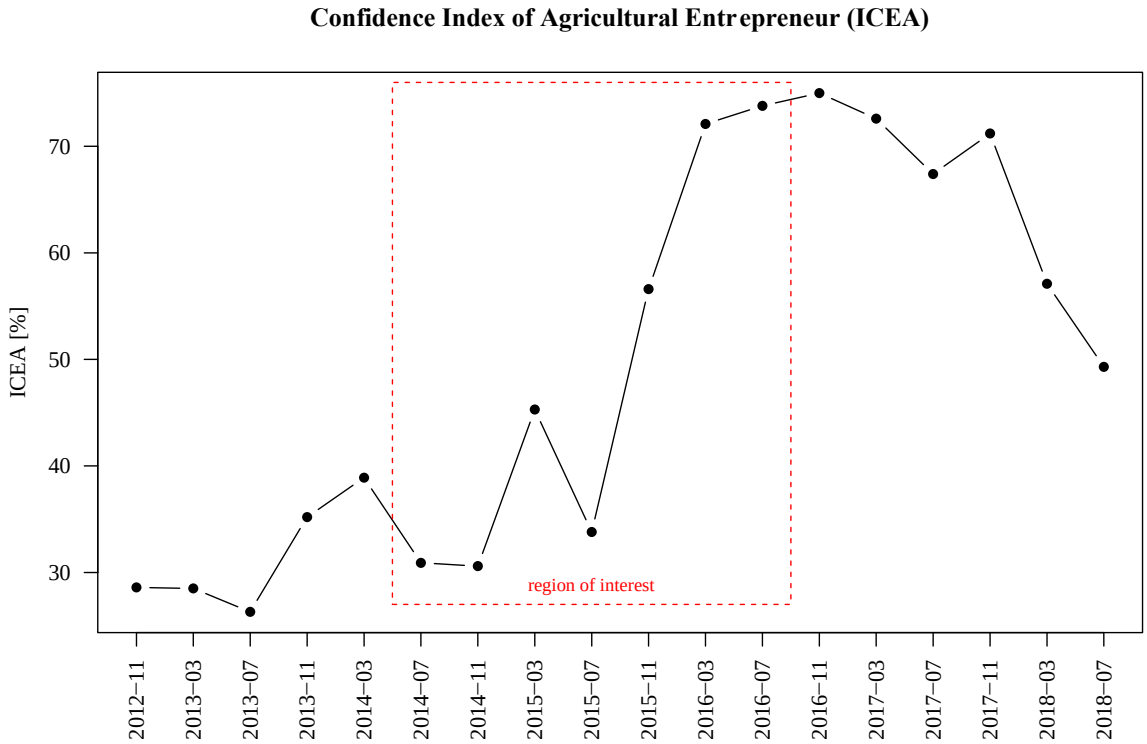

Figure 1. Confidence Index measured by survey.

To replicate the shape of the index evolution by mining Twitter, a set of Twitter users whose tweets contain the opinion needs to be outlined. CREA, the organization that conducts the survey, has an active Twitter account (@crea_arg) with more than 32K followers. Of that group of followers, not all are agricultural producers. Therefore, to filter users that are more similar to the survey responders, we selected users that follow, besides CREA account, four other accounts related to agriculture production concurrently: @intaargentina, @agrositio, @infocampoweb, and @ln_campo. With this consideration, we obtain a set with more than 7K relevant users. From that set, we randomly selected more than 2.5K users, to deal with the difficulty of obtaining old tweets from Twitter. All their tweets within the time-frame of interest were downloaded using a python script [20], obtaining more than 176.5K tweets.

2.2 Filtering relevant tweets

Not all tweets wrote by the users set previously discussed are related to macro-economic or sectorial conditions. To filter only those on-topic tweets, a list of words was manually constructed. We analyzed only tweets that contains at least one of those words. The list of words and their translation to English is shown in Table 1. The set of relevant tweets rises to almost 5K strings of texts with less than 280 characters.

Word	Translation	Word	Translation
gobierno	*government*	**política**	*politics*
ministerio	*ministry*	**economía**	*economy*
presidente	*president*	**económica**	*economic*
presidencia	*presidency*	**gobernador**	*governor (male)*
ministro	*minister (male)*	**gobernadora**	*governor (female)*
ministra	*minister (female)*	**gobernación**	*governance*
inflación	*inflation*	**situación**	*situation*

Table 1. List of relevant words

2.3 Obtaining sentiment aggregated score

To obtain the sentiment of filtered tweets, the Google Cloud, Natural Language API [10] was used. The

tool returned one sentiment score between 0 and 1, representing overall opinion, feeling, or attitude expressed in each tweet. First, the score was averaged by month and user, and then the average was aggregated by month among all users.

3 Results

Obtained Twitter sentiment mean score by month compared to the survey results shows a noteworthy parallelism. See Figure 2.

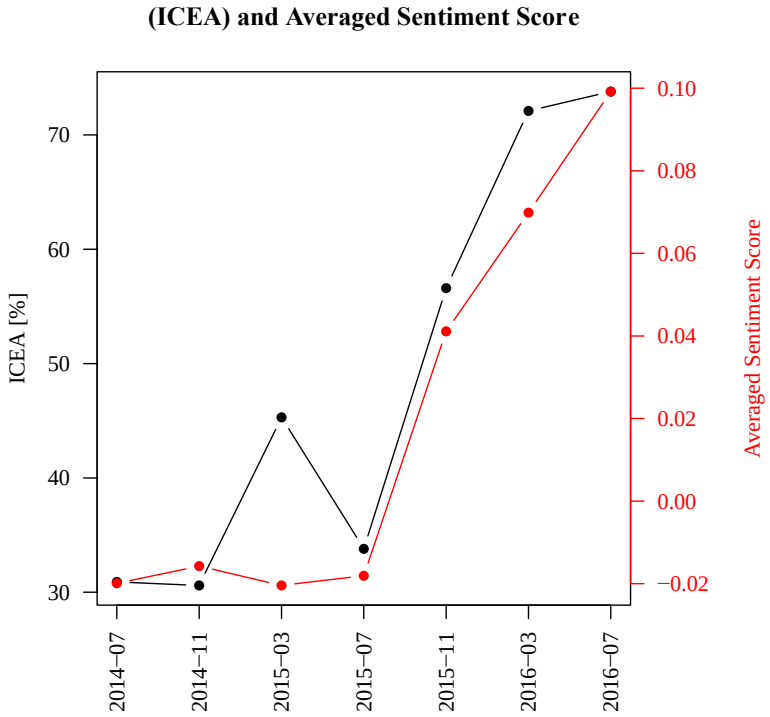

Figure 2. ICEA vs. Averaged Sentiment Score

4 Discussion

To verify the results, some tweets scored positive and negative are discussed (see Figure 3).

For instance, Maria Paula's tweet says: *"Small milk producers closing their farms, the situation is getting worse"*. This tweet was scored -0.89. Scored with -0.89 as well, Alejandro Grané's tweet claims that: *"Agricultural sector is in its worst time ever. No response from a leaving government with uncertainty about what is about to come"*, implying a very low confidence about the future. On the positive side, we have the Livestock Secretary of the Province of Córdoba account tweeting *"Governor congratulated the organization and acknowledged workers, producers and dairy industry"* with a score of 0.9. Positive tweet from Flavio Luetto Sainz says *"This is excellent news for the Argentine economy"*. Both tweets imply a positive confidence about present and future conditions.

This experiment shows evidence of the hypothesis that mining social media can provide relevant information to estimate the confidence of a group of people.

5 Acknowledgment

This research was funded by the Peruilh PhD scholarship and the UBACyT program (project code 20620170200003BA).

Figure 3. Tweets scored positive and negative.

References

[1] https://techcrunch.com/2016/11/10/social-media-did-a-better-job-at-predicting-trumps-win-than-the-polls/

[2] Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. Icwsm, 10(1), 178-185.

[3] Bae, J. H., Son, J. E., & Song, M. (2013). Analysis of twitter for 2012 South Korea presidential election by text mining techniques. Journal of Intelligence and Information Systems, 19(3), 141-156.

[4] Larsson, A. O., & Moe, H. (2012). Studying political microblogging: Twitter users in the 2010 Swedish election campaign. New Media & Society, 14(5), 729-747.

[5] Sisco, M. R., Bosetti, V., & Weber, E. U. (2017). When do extreme weather events generate attention to climate change?. Climatic Change, 143(1-2), 227-241.

[6] Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock mar-ket. Journal of computational science, 2(1), 1-8.

[7] O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. A. (2010). From tweets to polls: Linking text sentiment to public opinion time series. IC-WSM, 11(122-129), 1-2.

[8] Mostafa, M. M. (2013). More than words: Social networks' text mining for consumer brand sentiments. Expert Systems with Applications, 40(10), 4241-4251.

[9] https://github.com/Jefferson-Henrique/GetOldTweets-python

[10] https://cloud.google.com/natural-language/