

## Social Media Analytics

### Challenge

Volumes of timely, diverse, and multilingual data from social media have been successfully used to answer social science, political science, computational linguistics, and sociolinguistics questions. These include detecting real-world events, analyzing health on a large scale, detecting emotions, mining opinions, and predicting user-centric analytics. Existing models for classification and regression fail to generalize and do not take into account the dynamic and multilingual aspects of social media. While older models rely on textual content and other metadata our new models add image content into the mix.

### Approach

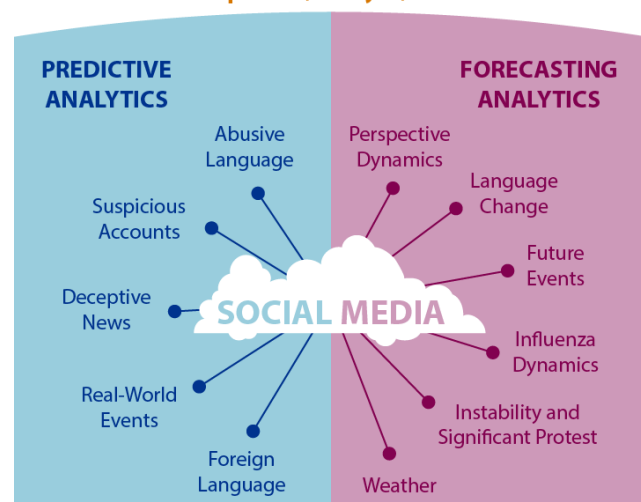
Our approach relies on developing neural network architectures capable of forecasting and nowcasting from social media signals including text, social graphs, and images.

Our predictive models

- » Discover deceptive news across languages
- » Detect abusive content

We develop predictive and forecasting analytics to advance understanding, analysis, and effective reasoning about extreme volumes of dynamic, multilingual, and diverse real-world social media data.

### Models to predict, analyze, and forecast

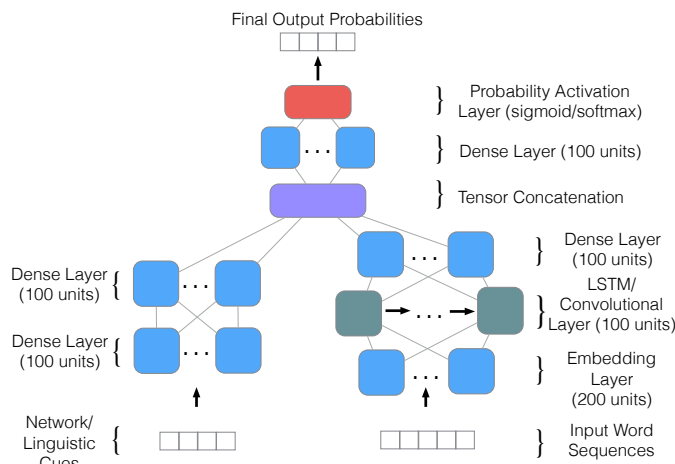


Social media predictive and forecasting analytics capabilities.

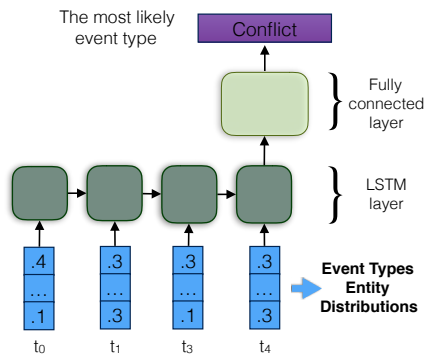
- » Classify suspicious accounts to be suspended or deleted
- » Detect real-world events
- » Predict a user's native language from English posts.

### Our forecasting models

- » Anticipate geo-spatial perspective dynamics for 10+ languages
- » Detect changes in language during crisis events
- » Anticipate future events, instability, significant protest, influenza dynamics, and weather across geolocations.



Model architecture for predicting types of deceptive news posts – propaganda, hoaxes, clickbait and satire.



Model architecture for forecasting future events e.g., conflict, nature, politics, business etc. from entity and event distributions in social media.

## Methodology

Our neural network architectures rely on single language and cross-language text representations such as words, characters, and bytes. It uses Recurrent Neural Networks such as Long Short-Term Memory and Gated Recurrent Units as well as Convolutional Neural Network models. We take advantage of

1. The “late-fusion” technique to incorporate different types of signals into our models and learn the combined representation of multiple input streams such as images and text, networks and linguistic cues, etc.
2. The ability to incorporate external knowledge for tasks with limited annotated data.

Our models take advantage of various social media signals including language, communication behavior, social network structure, events and entities, topics, syntactic and stylistic markers, and visual content.

Our predictive and forecasting analytics demonstrate how machine learning and deep learning models can be utilized across many tasks, domains, and languages.

## Impact

Our social media predictive analytics will

- » Enhance understanding of online give and take during a crisis
- » Detect and prevent the spread of abusive content
- » Identify and eliminate bots and trolls in social networks.

Moreover, predicting deceptive news and understanding deceptive content's strength and polarity, as well as forecasting dynamic perspectives jointly across locations and languages may enable analysts to defend against not only what has happened but what is perceived to have happened. Finally, the ability to forecast weather, disease dynamics, and world events like protests and civil unrest can help governments make plans to safeguard citizens and visitors from potential casualties and monitor essential supplies.

### CONTACT

**Svitlana Volkova**  
Senior Research Scientist  
(509) 372-6585  
svitlana.volkova@pnnl.gov

**Mark Greaves**  
Initiative Lead  
(206) 528-3300  
mark.greaves@pnnl.gov



Proudly Operated by **Battelle** Since 1965