EXCEL Homework 1

Search the web for some predictive algorithm currently or recently being used. You can use either the NYC Illegal Conversions, or the Chicago Gang Violence algorithm, of the as examples of what to look for.

**Describe the sources of the projects data. Describe the problems they had with using the data. Finally describe how they solved those problems.**

This is a project designed to predict civil unrest. Apparently it’s currently in use analyzing events in Latin America and is being modified for use in the Middle East.

**Sources**

Activist blogs

Newsfeeds (medical, political)

Tweets

Economic indicators

Opinion polls

Wikipedia edits

Weather data

Google Flu Trends

Parking lot images

Online restaurant reservations

**Data Project Name:**

EMBERS (Early Model-Based Event Recognition using Surrogates)

**Describe the Algorithm:**

The premise of the algorithm is that population-level behavior changes precede events of interest. The algorithm attempts to identify significant behavior changes through the analysis of its data sources and use historical data to classify what events typically follow these types of behavior change.

**Links that describe results and its data sources (up to 3):**

<https://www.eng.vt.edu/news/researchers-study-new-ways-forecast-critical-societal-events>

<http://www.basistech.com/predictive-analytics-case-study/>

http://1jv37d2ck4sy10rozi23w06t.wpengine.netdna-cdn.com/wp-content/uploads/pdf/EMBERS-Case\_Study.pdf

**Provide one sentence descriptions to 3 problems they faced with data, and give a one sentence solution they found for each one.**

Unstructured text (tweets, newsfeeds, etc.) is messy and hard to analyze. The project used a commercial text analytics tool (Rosette) to aid in identifying, normalizing, and tagging data found in free text.

Different countries and cultures require different behavior models. The project needed to revise the data model for different cultures as the same behavior in one culture may lead to a different outcome in another.

The events they want to predict and the data sources are very diverse. The project decided to create multiple specialized algorithms to focus on different classes of events, and then later combine the specialized analysis into the final result.

**Provide a one sentence example of why or why it did not work or is or is not working overall.**

According to the case study listed above, after two years the project met or exceeded all but one of the performance criteria set at the start of the project.