**Possible Model Features**

|  |  |  |
| --- | --- | --- |
| Feature | Why relevant? | Data source? |
| Past ILINet values | Current trends likely to be similar in some way to previous trends | CDC FluView |
| Virologic test results of season | Intensity patterns could be related to what type of virus is predominantly circulating | CDC FluView |
| Measure of backfill/uncertainty | Initial published values on FluView don’t always match up with final version, which is what models are evaluated on | Delphi Epidata |
| Week of season | Whether ILINet is going up or down from previous week is likely a function of when in the season you are | MMWR calendar |
| Social media data | Mentions of flu or symptoms on social media could be near-real time indicator of flu activity | Healthtweets,  Wikipedia? – how to get?  Google Trends |
|  |  |  |

**Google Trends:**

Get search terms from Google Correlate…

|  |  |
| --- | --- |
| Search term | Value |
|  |  |
|  |  |
|  |  |
|  |  |

Could get Google Trend data by state then weight by relative state population

Google Trend data gets noisy at state level, especially in smaller states – potentially take biggest state in each region and use that as a proxy??

**Possible modeling approaches**

Combine multiple models into one – uniform model, historical percentages, ARIMA, nearest neighbor approach,

Model each HHS region separately, then weighted average into overall US prediction

Model each age group separately? This could be esp effective using virologic data

* Check how US data compares to regional data – probably just linear combination so can come up with weights from beta coefficients from simple linear model

**Possible models to include**

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Data needed | How to make | Status |
| Weighted historical average | ILINet, virologic data | Create historical average models for H1 and H3 seasons.  For each week, weight H1/H3 null models by cumulative % of samples testing for each |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**Data source notes**

Google trends data can be accessed using gtrendsR package. Values are relative to max value in period specified – i.e. that gets a value of 100 and everything else is adjusted accordingly. As a result, values are \*highly\* context dependent based on dates. If data are pulled season by season, couldn’t really use to make inferences on