Named storms are typically predicted based on past occurrences and current measures of factors in the climate.

There have been great strides forward made in the science of forecasting hurricanes, but there is still a lot to do. One major problem is accuracy. The National Hurricane Center has been forecasting the paths of hurricanes since the early 1950’s. They issue 120 hour, 96 hour, 72 hour, 48 hour, 24 hour, and 12 hour forecasts. (The 120 hour and 96 hour forecasts were introduced in 2003.) The error decreases as the time before landfall decreases. The error has also decreased over the years as models become more accurate (NOAA, 2004). Despite becoming more accurate, the error is still relatively large.

How to Identify the Most Important Predictor Variables in Regression Models

<http://blog.minitab.com/blog/adventures-in-statistics-2/how-to-identify-the-most-important-predictor-variables-in-regression-models>

Table 1: Predict the last week of the months

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Training data | Test data | MSE | Training data | Test data | MSE |
| 01011016 - 24012016 | 25012016 - 31012016 | 0.02703264 |  |  |  |
| 01022016 - 22022016 | 23022016 - 29022016 | 0.02670071 | 01012016 - 22022016 | 23022016 - 29022016 | 0.03297402 |
| 01032016 - 24032016 | 25032016 - 31032016 | 0.02336329 | 01012016 - 24032016 | 25032016 - 31032016 | 0.02946526 |
| 01042016 - 23042016 | 24042016 - 30042016 | 0.03535662 | 01012016 - 23042016 | 24042016 - 30042016 | 0.03149145 |
| 01052016 - 24052016 | 25052016 - 31052016 | 0.02169727 | 01012016 - 24052016 | 25052016 - 31052016 | 0.02962781 |
| 01062016 - 23062016 | 24062016 - 30062016 | 0.02361385 | 01012016 - 23062016 | 24062016 - 30062016 | 0.02803105 |
| 01072016 - 24072016 | 25072016 – 31072016 | 0.02584931 | 01012016 - 24072016 | 25072016 – 31072016 | 0.02758849 |
| 01082016 - 24082016 | 25082016 – 31082016 | 0.02141876 | 01012016 - 24082016 | 25082016 – 31082016 | 0.02678308 |
| 01092016 - 23092016 | 24092016 - 30092016 | 0.02387472 | 01012016 - 23092016 | 24092016 - 30092016 | 0.02633638 |
| 01102016 - 24102016 | 25102016 – 31102016 | 0.01863817 | 01012016 - 24102016 | 25102016 – 31102016 | 0.026007 |
| 01112016 - 23112016 | 24112016 - 30112016 | 0.01236498 | 01012016 - 23112016 | 24112016 - 30112016 | 0.02557195 |
| 01122016 - 24122016 | 25122016 – 31122016 | 0.04776085 | 01012016 - 24122016 | 25122016 – 31122016 | 0.02597173 |
| Average | | 0.0256393 |  |  | 0.028168 |

Table 2: Predict monthly data

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Training data | Test data | MSE | Training data | Test data | MSE |
| Jan 2016 | Feb 2016 | 0.03943264 |  |  |  |
| Feb 2016 | Mar 2016 | 0.02966795 | 012016 - 022016 | 032016 | 0.02927437 |
| Mar 2016 | Apr 2016 | 0.02774618 | 012016 -032016 | 042016 | 0.03204248 |
| Apr 2016 | May 2016 | 0.02276284 | 012016 - 042016 | 052016 | 0.02939354 |
| May 2016 | Jun 2016 | 0.02293082 | 012016 - 052016 | 062016 | 0.02823762 |
| Jun 2016 | Jul 2016 | 0.02302146 | 012016 - 062016 | 072016 | 0.02758651 |
| July 2016 | Aug 2016 | 0.02605833 | 012016 - 072016 | 082016 | 0.02664009 |
| Aug 2016 | Sep 2016 | 0.02263518 | 012016 - 082016 | 092019 | 0.02715848 |
| Sep 2016 | Oct 2016 | 0.01808771 | 012016 - 092016 | 102016 | 0.02600165 |
| Oct 2016 | Nov 2016 | 0.01950643 | 012016 - 102016 | 112016 | 0.02573006 |
| Nov 2016 | Dec 2016 | 0.03605422 | 012016 - 112016 | 122016 | 0.02619549 |
| Average | | 0.026173 | Average | | 0.02829 |

Data 3: Predict 30% of the whole zillow’s 2016 training data (920275 datasets)

model1\_a = lm(logerror ~ bathroomcnt + bedroomcnt + fips + latitude + longitude + propertylandusetypeid + roomcnt + taxvaluedollarcnt + calculatedfinishedsquarefeet + yearbuilt, data = train1)

training data 70% of the whole zillow’s 2016 training data (920275 datasets)

testing data 30% of the whole zillow’s 2016 training data (920275 datasets)

MSE = [1] 0.002548212