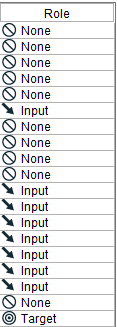
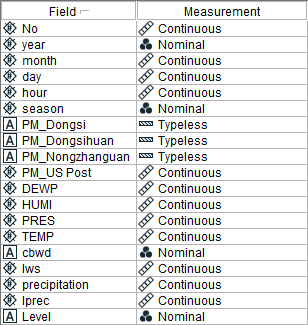
**The reasons that affect pm2.5 in City**

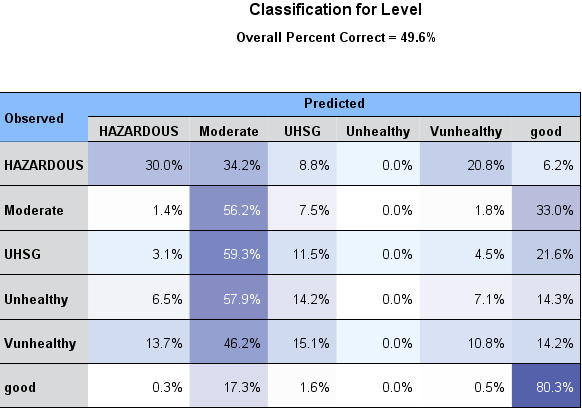
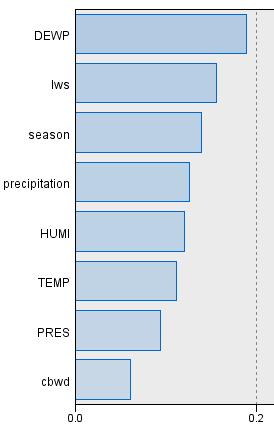
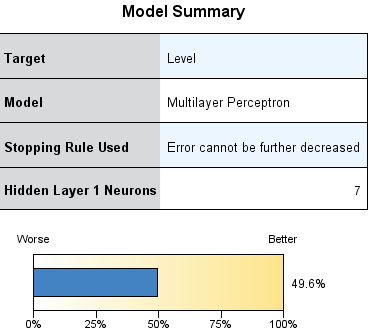
PM 2.5 refers to atmospheric particulate matter that has a diameter less than 2.5 micrometers. it floats invisibly through the air. but long-time inhalation of air with pm 2.5 might exacerbate health problems and cause potential illness for people. as the development of globalization, not only the city of Beijing, more and more people from around the world would face the menace of environmental pollution. Therefore, the object of the project is to analyze what factors could influence the PM index.



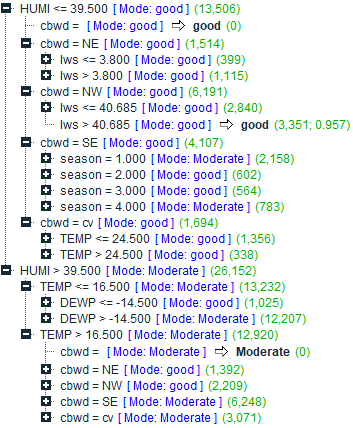
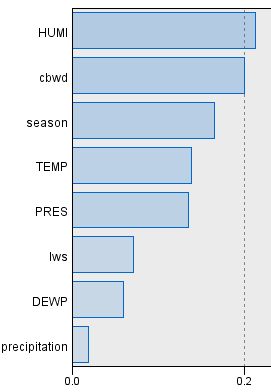
The dataset contains 50387 records, collected from Peking University. only 49579 data are valid by using data audit. Each row represents Peking’s hourly weather condition hour. The dataset includes eighteen attributes. We choose the US embassy recorded(US.post) to become our dependent variable. First, we use Excel to remove the row with empty data record from the table. Because the dataset includes two nominal type column(season and cbwd,’wind direction’),we could not doing the regression analysis. Thus we created a new target variable, Level, to classify the data from our target variable depend on the range of PM 2.5 index. 0 to 50: Good. 51 to 100: moderate. 101 to 150: unhealthy for sensitive group. 151 to 200: unhealth. 201 to 300: very unhealthy. over 300: hazarouds.

About the role of attribute, we removes all four attributes about time from the input, because these records are not suitable(for example,irpre is accumulated hourly precipitation) or irrelevant to the object of the project. three attributes about location was removed because these records have large numbers of empty record that would influence the analysis result. and one location attribute was replaced by the new variable we create. besides, we removed an attribute - Cumulated. because the data feature accumulated and collected from twenty-four hours, other attribute is independent. so the attribute will influence the outcome of analysis as well. Finally, we choose eight attributes that are relevant to the object of the project from the dataset.

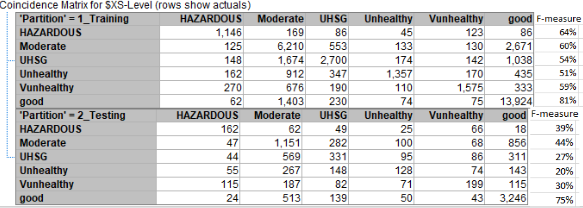
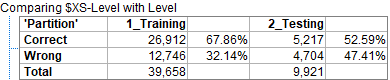
Due to around 50000 data, we use 80% data for data training, 20% for data testing, and use two algorithm models, which are ANN and DT, to analyze the data source.



**1. ANN Model,** we can see that the accuracy of the model is 49.6%. In the predicted importance, temperature is the most significant influence, then followed by DEWP, LWS,season , PRES and other factors. In Classification for level we can see that “Good” class has the highest accuracy(80.3%) , next is moderate (56.2%), and Hazardous(30%).



**2. Decision Tree Model**, among all predictor importance factors, Humidity is the most significant influence, then followed by combined direction of winds, season and other factors. When in a situation that humidity equal or under 39.5, combined direction of wind are north-west and wind speed over 40.68 the level has 95% accuracy to be “good”. Look at the whole decision tree, we find that when Humidity is no more than 39.5 there the model has higher accuracy, when the Humidity over 39.5 it was hard to have clear outcomes.



**4. Ensemble:** We use the Ensemble to combine two model’s analysis outcome and we found Training accuracy has reached to 67.86% and testing accuracy has reached to 52.59%. In Coincidence matrix table we can see that “good” class has highest accuracy which is reached to 75% in testing and 81% in training. The Unhealthy for sensitive group has the lowest F-measure which is 54% in Training and 27% in Testing. Same as Unhealthy(51% training, 20% testing ) and Very Unhealthy class(59% training, 30% testing). The moderate class has 60% f-measure in training and 44% in testing. The Hazardous has 64% f-measure in training and 39% in testing.

Model result: Decision Tree is better model in this case. When Humidity over 39.5 it may disturb the model to get clear outcome, when wind speed over 40.68 also has good influence to the Beijing’s PM2.5.

**Conclusion**

After researching, we integrated the two models and combined with our research results we found that the HUMI,CBWD, TEMP and Season had the most significant influence on pm2.5. Based on the research we found, in high humidity environment (Normally over 60%), which always has mist in the air, the drop of the water in the air and PM2.5 will combine with each other and become the fog and haze.In this situation The measurement equipment normally cannot distinguish PM2.5 and water drop when combined with each other, due to PM2.5 and mist’s diameter both less than 2.5 micrometers. These problems cause the result of artificially high value of (U.Spost)pm.2.5. The wind from North-west in Beijing is from dry areas which also links to the Beijing’s humidity level. For the “season”, based on our research, When Beijing in Autumn(3), it always in windy weather and the direction of the wind normally is from North, which also link with the Humidity and wind speed.

Thus we have reason to say that the humidity and wind speed are the two most important environment elements to effect the detected value of pm2.5. and wind speed is the most important and direct element to in our input to effect the actual value of pm2.5

In reality, researching this topic that enables us to know what factors are responsible for the strong pm2.5. What’s more, studying this area can deepen our understanding of relation between urban climate and PM2.5’s detection value. High occurrence of extreme haze episodes in recent years not only leads to a global concern due to its adverse health effects but also triggers the Chinese government to tackle the severe air quality problem, especially PM2.5 pollution.

dataset:pm2.5

[http://archive.ics.uci.edu/ml/datasets/PM2.5+Data+of+Five+Chinese+Cities#](http://archive.ics.uci.edu/ml/datasets/PM2.5+Data+of+Five+Chinese+Cities)

# The Effects of PM2.5 Concentrations and Relative Humidity on Atmospheric Visibility in Beijing,01/28/2019,[Xiaoyan Wang](https://agupubs.onlinelibrary.wiley.com/action/doSearch?ContribAuthorStored=Wang%2C+Xiaoyan) [Renhe Zhang](https://agupubs.onlinelibrary.wiley.com/action/doSearch?ContribAuthorStored=Zhang%2C+Renhe) [Wei Yu](https://agupubs.onlinelibrary.wiley.com/action/doSearch?ContribAuthorStored=Yu%2C+Wei)

<https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018JD029269>