

DS 670 – Assignment 13 - Final Manuscript

Aakash Parwani

April 22, 2017

- **Abstract:**

The inspiration driving this wander is to play out an examination of atmosphere data. Here, examination portrays point by point examination of fundamental atmosphere data parameters for performing atmosphere foreseeing. Preliminary data will be assembled from CityPulse an online store that offers different semantically cleared up datasets accumulated from associates of the CityPulse EU FP7 broaden and relevant resources for splendid city data. The hypothesis of the survey is delivered as H1: Today a ton of imperative data and sensor information remain unused or are limited to specific application spaces due to incalculable progressions and setups (atmosphere takes note). In this manner, a get-together of basic information from various sources is done physically and as a general rule, it is out-dated. A t-test with the expectation of complimentary examples is used to address the hypothesis. The discussion of the survey researched extraordinary information on atmosphere database. Still, there are basic calculates that are truant this survey which can be considered in future for arranging powerful methodology. The eventual outcome of the survey is a productive and reasonable representation of atmosphere information database and traverse region points of confinement to engage honestly to goodness gagging and atmosphere forewarning reports that reinforce the coordination and examination of heterogeneous data and information sources and empowers the progression of innovative consistent smart city applications.

- **Work by Competitors:**

Network Architecture for Small X-BAND Weather Radars (Pedersen, 2007)

In recent years the use of small inexpensive X-band radars for meteorological and hydrological purposes has increased significantly. Compared to the traditional C-band and S-band radars the X-band weather radar has the advantage of high temporal and spatial resolution and low financial cost; however, the trade off is attenuation due to X-band technology and short range due to the higher spatial resolution.

In relation to quantitative precipitation estimation (QPE) and forecasting (QPF) it is necessary to address the attenuation issue. Over the next two years a test bed with five X-band radars of the type Local Area Weather Radar (LAWR) will be set up south of Aarhus in Denmark. The three radars will be placed so that they overlook the same area and thereby are capable of intercepting a precipitation event from different angles, altitudes and ranges.

The overall aim is to construct a network of small X-band radars which are capable of automatic inter-calibrating and inter-correcting for attenuation on the fly by online communication. From the beginning the LAWR has been equipped with a set of attenuation corrections algorithms. However, verification of these algorithms has not been possible so far, due to lack of multi radar measurements of rainfall events. The prospects of verifying and possibly improving the method used for correction of attenuation using radar networks are promising. The results from the test bed are expected to optimize attenuation handling and improve now casting capabilities.

Average Daily Air Temperature's Long-Range Forecast Using Inductive Modeling (Zubov, 2013)

DS 670 – Assignment 13 - Final Manuscript

In this paper, long-range forecasting average daily air temperature using inductive method was proposed. Principle of high-impact weather events substantiates the different places' interaction by atmosphere, hydrosphere, landmass, biosphere, etc. Forecasting model reasoning's first stage is selection of three most data-related places using Pearson product-moment correlation coefficient, which has to be greater than 0.8 in absolute value. 66 datasets were acquired from NOAA Satellite and Information Service. Second stage is finding weighting coefficients of forecasting model and criterion "minimum of regularity plus maximum of conjunctions" by combinatorial algorithm. This concept is illustrated by Skopje Airport's forecasting model and criterion reasoning, which include datasets from Beijing (China), Ulaanbaatar (Mongolia), and Paphos Airport (Cyprus). Results (conjunctions' percent is 74.8 %, mean absolute error (MAE) is up to 5.7 °F, 166 days lead-time) showed an efficiency of proposed approach. Similar results were achieved for Kiev (MAE is up to 7.2 °F, 167 days lead-time) and Washington National Airport (MAE is up to 6.07 °F, 173 days lead-time). Web-site prototype www.weatherforecast.tk was developed using Ms Windows Azure public cloud computing technology. Proposed approach is characterized by high accuracy, final linear difference equations' simplicity, low computational complexity, and user-friendly interface, which is very suitable for meteorological services.

Developing Cloud Computing Novel Computational Methods for Improving Long-Term Weather Global Forecast (Dmytro, 2012)

Weather data mining methods and forecast algorithms have been of long standing interest. Recent research based on the global satellite data and special synergetic methods showed possibility of the long-term (up to half a year ahead) forecast with up to 10 % average mistake (standard is 20 %). Particularly, the average daily air temperature forecast's mistake is up to 6.5 % for Skopje Airport (half a year ahead). This approach is characterized by the final linear difference equations' simplicity and the high computational complexity of the above equations reasoning. The cloud computing web-site's prototype was developed (weatherforecast.tk). Main research proposals: improving the user interface based on 3D or/and ubiquitous computing technologies; developing new synergetic methods for the appropriate realization in the multithread cloud application, including the code and data parallelization; increase of the forecast parameters' quantity (e.g., precipitation). This paper main results are: precipitation's long-term (up to half a year ahead) forecast has very low quality now, and, therefore, it is not recommended for practice; the forecasting places' quantity is changed modifying the text file in the cloud application's package; the web-site <http://weatherforecast.tk> user interface was enhanced using 3D Chart diagram.

• **Contribution:**

Let us quickly comprehend the normal commitment of our investigation for the health of atmosphere. We will go through importance of each weather analysis component considered for this project and understand the individual contribution on the health of atmosphere.

Dew Point (In Degree Celsius): - Dew point is the temperature at which airborne water vapor will amass to shape fluid dew. A higher dew point proposes there will be more steepness discernible all around. Dew point is every so often called ice minute that the temperature is underneath nippy. The estimation of dew indicates is connected mugginess.

DS 670 – Assignment 13 - Final Manuscript

Real-Time application, most builders appreciate that development can shape when warm, sodden air encounters a cool surface. The development is shocking, and makers need to keep up a key separation from it. There's an answer, be that as it may: According to building specialists, we can expect development issues in dividers by choosing a divider's temperature profile and playing out a dew-point tally. It's certainly profitable to know whether you're sheathing will be over the dew point or underneath the dew point in winter. When sheathing is underneath the dew point, it's most likely going to gather soggy. Warm sheathing is better than anything cold sheathing.

Unfortunately, though, temperature profiles and dew-point calculations have been misunderstood and misused for years. In the excellent book, *Water in Buildings*, William Rose wrote, "The language 'reaching dew point' seems to indicate that one could plot a temperature profile through a wall, find the point where that profile intersects a horizontal line indicating indoor dew point temperature, and expect burgeoning water at that location. This impression is decidedly incorrect. If water accumulates, it does on the surfaces of materials, not within the thickness of materials."

Humidity (In Percentage): - Humidity is the measure of water vapor noticeable all around. Water vapor is the vaporous condition of water and is imperceptible. Humidity shows the probability of precipitation, dew, or haze. Higher dampness diminishes the viability of sweating in cooling the body by decreasing the rate of dissipation of dampness from the skin. There are three primary estimations of humidity: total, relative and particular.

Relative humidity is the most frequently encountered measurement of humidity because it is regularly used in weather forecasts. It's an important part of weather reports because it indicates the likelihood of precipitation, dew, or fog. Higher relative humidity also makes it feel hotter outside in the summer because it reduces the effectiveness of sweating to cool the body by preventing the evaporation of perspiration from the skin.

Humidity expects a key part in our step by step atmosphere. Without water vapor detectable all around, our atmosphere may take after the atmosphere on Mars. Would you have the capacity to imagine presence without fogs, rain, snow, thunder, or lightning?

So how does humidity impact us on a hot day? Individuals are sensitive to changes in humidity, in light of the way that our skin uses the air around us to discard clamminess as sweat. In case the relative humidity is high, the air is starting at now drenched with water vapor and our sweat won't vanish. Right when this happens, we feel more smoking than the genuine temperature.

Similarly, low humidity can make us feel cooler than the genuine temperature. This happens in light of the fact that the dry air sweats dissipate more rapidly than expected.

Pressure (In mBar): - Atmospheric pressure, here and there additionally called barometric weight is the weight applied by the heaviness of air in the environment of Earth (or that of another planet). Much of the time climatic weight is nearly approximated by the hydrostatic weight created by the heaviness of air over the estimation point.

DS 670 – Assignment 13 - Final Manuscript

The modification in air pressure after some time has basic suspecting recommendations. As pressure brings down after some time, especially if it is snappy, that implies that a low-pressure system or front is moving closer. This cutting down pressure exhibits an extending likeliness of precipitation. If the air pressure rises basically or stays well superior to anything expected for a drawn out extend of time, which is a sign precipitation is more stunning.

Air pressure at a point is the compel per unit zone applied by the heaviness of air above it. This implies air pressure differs extensively with stature, diminishing with expanding rise. To take into consideration correlation between estimations at various areas, all pressure readings are changed over to mean ocean level pressure. This implies the elevation of every station measuring pressure must be known unequivocally. The unit of pressure is the pascal (Pa), as this is a little unit of pressure, in meteorological practice the hectopascal (hPa) is utilized (1hPa=100Pa). Lines of equivalent pressure drawn on a climate guide are called isobars. And being an essential atmosphere variable, pressure is an imperative variable in portraying everyday climate variety and in climate anticipating and displaying.

Temperature (In Degree Celsius): - Temperature is a level of hotness or coldness the can be measured utilizing a thermometer. It's likewise a measure of how quick the particles and atoms of a substance are moving. Temperature is measured in degrees on the Fahrenheit, Celsius, and Kelvin scales.

Guesses in light of temperature and precipitation are fundamental to cultivation, and in this way to dealers inside product markets. Temperatures figures are used by administration associations to gage ask for over coming days. On a standard commence, people use atmosphere evaluations to make sense of what to wear on a given day.

Power and gas associations rely on upon atmosphere assessments to speculate ask for which can be immovably affected by the atmosphere. They use the sum named the degree day to choose how strong of a use there will be for (warming degree day) or (cooling degree day). These sums rely on upon a consistently typical temperature of 65 °F (18 °C). Cooler temperatures drive warming degree days (one for each degree Fahrenheit), while more blazing temperatures oblige cooling degree days. In winter, genuine crisp atmosphere can achieve a surge well known as people turn up their warming. Correspondingly, in summer a surge looked for after can be associated with the extended use of air trim structures in a hot atmosphere. By retribution a surge looked for after, administration associations can purchase additional provisions of constrain or regular gas before the cost increases, or in a couple conditions, supplies are restricted utilizing brownouts and power blackouts.

Wind Direction (In Degrees): - Wind direction is represented by the course from which it starts. For example, a northerly wind blows from the north toward the south. The wind bearing will vitally affect the ordinary atmosphere. You can frequently be given a twist course and you will have a completely brilliant considered how the atmosphere will change and what atmosphere can be typical with that wind heading.

DS 670 – Assignment 13 - Final Manuscript

The wind heading will critically affect the ordinary atmosphere. You can as often as possible be given a wind heading and you will have a truly savvy considered how the atmosphere will change and what atmosphere can be ordinary with that wind bearing.

The average wind bearing that a range has for a particular time is known as the general wind. Right, when the curve is from the regular bearing then the atmosphere is generally common. Exactly when the twist moves a long way from the all-encompassing course then it consistently demonstrates atypical or advancing atmosphere.

Wind direction changes routinely run with changes in the atmosphere. The wind streams cyclonically around low-weight systems. If the wind modifies course in a cyclonic way it oftentimes infers a low weight or front is affecting the figure zone. A wind moving from the south routinely infers more smoking air is moving ever closer contort from the north as often as possible suggests cooler air is moving closer. Right, when the wind changes surprisingly it could be a frontal segment or wind direction change made by tempest surge.

Wind Speed (In Kmph): - Wind speed, or wind stream speed, is a basic climatic amount. Wind speed is brought on via air moving from high weight to low weight, more often than not because of changes in temperature. Wind speed influences climate determining, airplane and oceanic operations, development undertakings, development and digestion system rate of many plant species, and endless different ramifications.

The wind speed will accept a key part in the surface temperature in conditions where there is a strong temperature change with stature in the farthest point layer. In particular, a strong temperature sneaks past rate in the bit of the farthest point layer nearest the surface. The earth is warmed and cooled beginning from the most punctual stage. The wind mixes this air at ground level with air higher overtops. In the midst of the day when winds are light and the skies are clear, warmth will work at the surface. The temperature for this circumstance will tend to be sultrier than if the wind speeds were more grounded. This is in light of the fact that more grounded winds will mix the warm air near the surface with cooler air overhead.

On a fresh night, the opposite is the circumstance. Light winds amid the night will allow cool air to work at the surface. In case winds are more grounded than expected then the surface temperature will be more smoking since the wind will mix more sizzling air high up with the shallow cool air working at the surface. The wind speed is moreover basic in choosing the rate at which warm move in climate conditions will happen.

- **Data:**

CityPulse gives a game plan of open-source fragments and demonstrators for sharp city application engineers, offering access to and organization of the re-usable parts like: - Online-Data-Repository, CityPulse-3D-Map, City-Pulse-City-Dashboard, Data-Quality-Explorer, Social-Media-Analyzer.

The weather data for the city of Aarhus in Denmark is public for analyses purpose, available at [Weather Data](#) . The dataset is a collection of weather observations from the city of Aarhus. Measurements are recorded from February 2014 – June 2014 and August 2014 – September 2014.

DS 670 – Assignment 13 - Final Manuscript

Weather data values will be analyzed on the basis of components like - Dew Point, Humidity, Pressure, Temperature, Wind Direction and Wind Speed.

Data Collection & Processing

As we know that CityPulse provides online repository of weather data set for data analysis and exploration purpose. Hence, data collection will be done from that repository.

Data provided in online repository is in [.Json](#) file format. And in this project we will consider processing of only [.Csv](#) file format. So, we will perform transformation of .Json to .Csv file format.

For data analyses and exploration we will use **Apache Zeppelin** environment more specifically **Spark Module**.

Data Variables

For weather analysis we are considering six important variables Dew Point, Humidity, Pressure, Temperature, Wind Direction and Wind Speed. Let us explore the meaning of each variable in detail.

Date Time: - Date and time of weather observation. **Data structure**: YYYY/MM/DD HH:MM:SS

Dew Point (In Degree Celsius): - Dew point is the temperature at which airborne water vapor will accumulate to shape liquid dew. A higher dew point suggests there will be more soddenness observable all around. Dew point is now and again called ice minute that the temperature is underneath chilly. The measurement of dew point is related to humidity.

Data Type: Integer.

Humidity (In Percentage): - Humidity is the measure of water vapor noticeable all around. Water vapor is the vaporous condition of water and is imperceptible. Humidity shows the probability of precipitation, dew, or haze. Higher dampness diminishes the viability of sweating in cooling the body by decreasing the rate of dissipation of dampness from the skin. There are three primary estimations of humidity: total, relative and particular.

Data Type: Integer.

Pressure (In mBar): - Atmospheric pressure, here and there additionally called barometric weight is the weight applied by the heaviness of air in the environment of Earth (or that of another planet). Much of the time climatic weight is nearly approximated by the hydrostatic weight created by the heaviness of air over the estimation point.

Data Type: Integer.

Temperature (In Degree Celsius): - Temperature is a level of hotness or coldness the can be measured utilizing a thermometer. It's likewise a measure of how quick the particles and atoms of a substance are moving. Temperature is measured in degrees on the Fahrenheit, Celsius, and Kelvin scales.

Data Type: Integer.

DS 670 – Assignment 13 - Final Manuscript

Wind Direction (In Degrees): - Wind direction is represented by the course from which it starts. For example, a northerly wind blows from the north toward the south. The wind bearing will vitally affect the ordinary atmosphere. You can frequently be given a twist course and you will have a completely brilliant considered how the atmosphere will change and what atmosphere can be typical with that wind heading.

Data Type: Integer.

Wind Speed (In Kmph): - Wind speed, or wind stream speed, is a basic climatic amount. Wind speed is brought on via air moving from high weight to low weight, more often than not because of changes in temperature. Wind speed influences climate determining, airplane and oceanic operations, development undertakings, development and digestion system rate of many plant species, and endless different ramifications.

Data Type: Double.

• **Method:**

In this section we will understand the complete process and importance of all the steps used in method.

DATA COLLECTION: - **Data collection** is the route toward get-together and measuring data on elements of energy, in a set up efficient form that engages one to answer communicated inquire about inquiries, test hypotheses, and survey comes about. The data collection portion of research is essential to all fields of study including physical and humanistic systems, humanities, business, et cetera. While methodologies vary via prepare, the emphasis on ensuring accurate and reasonable collection proceeds as some time recently.

The importance of ensuring accurate and appropriate data collection

Despite the field of study or inclination for portraying data (quantitative, subjective), amend data collection is basic to keeping up the uprightness of research. Both the choice of sensible data collection instruments (existing, balanced, or starting late made) and apparently portrayed out a direction for their benefit utilize lessen the probability of blunders event.

Consequences from improperly collected data include

- inability to answer research questions accurately
- inability to repeat and validate the study
- distorted findings resulting in wasted resources
- misleading other researchers to pursue fruitless avenues of investigation
- compromising decisions for public policy
- causing harm to human participants and animal subjects

While the level of impact from broken data collection may contrast via prepare and the method for examination, there is the likelihood to realize uneven naughtiness when these investigation results are used to reinforce open procedure proposition.

DS 670 – Assignment 13 - Final Manuscript

Issues related to maintaining integrity of data collection:

The primary rationale for preserving data integrity is to support the detection of errors in the data collection process, whether they are made intentionally (deliberate falsifications) or not (systematic or random errors).

Most, Craddick, Crawford, Redican, Rhodes, Rukenbrod, and Laws (2003) describe 'quality assurance' and 'quality control' as two approaches that can preserve data integrity and ensure the scientific **validity** of study results. Each approach is implemented at different points in the research timeline (Whitney, Lind, Wahl, 1998):

1. Quality assurance - activities that take place *before* data collection begins
2. Quality control - activities that take place *during* and *after* data collection

Data Source: -

CityPulse gives a game plan of open-source fragments and demonstrators for sharp city application engineers, offering access to and organization of the re-usable parts like: - Online-Data-Repository, CityPulse-3D-Map, City-Pulse-City-Dashboard, Data-Quality-Explorer, Social-Media-Analyzer.

The weather data for the city of Aarhus in Denmark is public for analyses purpose, available at [Weather Data](#) . The dataset is a collection of weather observations from the city of Aarhus. Measurements are recorded from February 2014 – June 2014 and August 2014 – September 2014. Weather data values will be analyzed on the basis of components like - Dew Point, Humidity, Pressure, Temperature, Wind Direction and Wind Speed.

For data analyses and exploration we will use **Apache Zeppelin** environment more specifically **Spark Module**.

DATA TRANSFORMATION: - Data provided in online repository is in [.Json](#) file format. And in this project we will consider processing of only [.Csv](#) file format. So, we will perform transformation of .Json to .Csv file format.

Data transformation is the path toward changing over data from one association (e.g. a database record, XML file, or Excel sheet) to another. Since data often lives in different territories and designs over the attempt, data transformation is essential to ensure data from one application or database is conceivable to various applications and databases, a fundamental part for applications blend.

In a run of the mill situation where data should be shared, data is extricated from the source application or data stockroom, changed into another arrangement, and after that stacked into the objective area. Extraction, transformation, and stacking (together known as ETL) are the focal procedures of data mix. Contingent upon the way of the coordination situation, data may be consolidated, amassed, improved, compressed, or sifted.

The underlying stride of data transformation is data mapping. Data mapping chooses the relationship between the data segments of two applications and sets up headings for how the data

DS 670 – Assignment 13 - Final Manuscript

from the source application is changed before it is stacked into the goal application. By the day's end, data mapping produces the fundamental metadata that is required before the certified data transformation happens.

DATA CLEANING: - Data cleaning is the route toward recognizing and helping (or clearing) decline or mistaken records from a record set, table, or database and implies perceiving divided, misguided, off base or unessential parts of the data and after that supplanting, changing, or deleting the muddled or coarse data. Data cleaning may be performed keenly with data wrangling instruments, or as cluster get ready through scripting.

In the wake of cleansing, a data set should be relentless with other practically identical data sets in the structure. The abnormalities perceived or cleared may have been at first brought on by customer entry bungles, by degradation in transmission or limit, or by different data word reference implications of similar components in different stores. Data cleansing differentiations from data endorsement in that endorsement always infers data is rejected from the system at entry and is performed at the period of area, instead of on bundles of data.

DATA SPLITTING: - The slightest troublesome course for us to comprehend the limit of a judicious model to perform on future data is to endeavor to copy this certainty. Notwithstanding the way that we can't genuinely get to the future before it happens we can hold some of our starting at now available data and view it just as was data from what's to come. For example, if we are predicting which Internet advancements a web page visitor will tap on; we may collect perceptive models using data from two days earlier and make figures for yesterday. This is truly a faultless proliferation with the preferred standpoint that we obviously unquestionably know which notices were tapped on, and by whom, yesterday. We can along these lines differentiate our desires and the outcomes that truly happened. In advancing endeavors and credit peril models, we as a general rule work with data identifying with a singular point in time (or interval in time, for instance, one week, one month, and one campaign). Such data is consistently suggested as cross-sectional. For such issues we routinely separate the available data into discrete sections self-assertively, developing our models on one of these portions and using the other for farsighted model examination and possibly exhibit refinement.

And in our case Aarhus city data set will be divided in 60-40 ratio. In which **60%** will be **training part** and **40%** of data will go in **testing bucket**.

PRINCIPAL COMPONENT ANALYSIS: - Principal components analysis is a procedure for identifying a smaller number of uncorrelated variables, called "principal components", from a large set of data. The goal of principal components analysis is to explain the maximum amount of variance with the fewest number of principal components. Principal components analysis is commonly used in the social sciences, market research, and other industries that use large data sets.

Principal components analysis is commonly used as one step in a series of analyses. You can use principal components analysis to reduce the number of variables and avoid multi co linearity, or when you have too many predictors relative to the number of observations.

Example

DS 670 – Assignment 13 - Final Manuscript

A consumer products company wants to analyze customer responses to several characteristics of a new shampoo: color, smell, texture, cleanliness, shine, volume, amount needed to lather, and price. They perform a principal components analysis to determine whether they can form a smaller number of uncorrelated variables that are easier to interpret and analyze. The results identify the following patterns:

- Color, smell, and texture form a "Shampoo quality" component.
- Cleanliness, shine, and volume form an "Effect on hair" component.
- Amount needed to lather and price form a "Value" component.

MODEL BUILDING: - Model Building—choosing predictors—is one of those skills in statistics that is difficult to teach. It's hard to lay out the steps, because at each step, you have to evaluate the situation and make decisions on the next step.

If you're running purely predictive models, and the relationships among the variables aren't the focus, it's much easier. Go ahead and run a stepwise regression model. Let the data give you the best prediction.

But if the point is to answer a research question that describes relationships, you're going to have to get your hands dirty.

It's easy to say "use theory" or "test your research question" but that ignores a lot of practical issues. Like the fact that you may have 10 different variables that all measure the same theoretical construct and it's not clear which one to use. Or that you could, theoretically, make the case for all 40 demographic control variables. But when you put them all in together, all of their coefficients become non significant.

All the variables in Aarhus city dataset are of type continuous. We will build **linear models** for prediction of response variable. Below are some linear models that we will build on our dataset.

- Simple Linear Regression
- Multiple Linear Regression
- Additive Model

MODEL EVALUATION & SELECTION: - Model Evaluation is an integral part of the model development process. It helps to find the best model that represents our data and how well the chosen model will work in the future. Evaluating model performance with the data used for training is not acceptable in data mining because it can easily generate overoptimistic and over fitted models. There are two methods of evaluating models in data mining, Hold-Out and Cross-Validation. To avoid over fitting, both methods use a test set (not seen by the model) to evaluate model performance.

Hold-Out

In this method, the mostly large dataset is randomly divided to three subsets:

- Training set is a subset of the dataset used to build predictive models.

DS 670 – Assignment 13 - Final Manuscript

- Validation set is a subset of the dataset used to assess the performance of model built in the training phase. It provides a test platform for fine tuning model's parameters and selecting the best-performing model. Not all modeling algorithms need a validation set.
- Test set or unseen examples are a subset of the dataset to assess the likely future performance of a model. If a model fit to the training set much better than it fits the test set, over fitting is probably the cause.

Cross-Validation

When only a limited amount of data is available, to achieve an unbiased estimate of the model performance we use k-fold cross-validation. In k-fold cross-validation, we divide the data into k subsets of equal size. We build models k times, each time leaving out one of the subsets from training and use it as the test set. If k equals the sample size, this is called "leave-one-out".

PREDICTION: - After selection of best model. We will use that model on **testing data set** for prediction of **Wind Speed**, which is our response variable.

• Results:

This section is divided into results obtained from various analysis performed at different stages of this project.

- **Facts obtained during data exploration**
 - a) **March** is the month of high humidity.
 - b) And day wise **Tuesday** is the highest humid day.
 - c) **April, June & September** are months with high temperature.
 - d) **Monday & Friday** are working days with high temperature.
- **Pearson correlation results**
 - a) Pearson correlation coefficient for **dew point & wind speed** was "0.0032" with P-Value of "0.715". This means there is positive relation between dew point and wind speed but this relationship is not that solid. And these results also convey that an increase in dew point will not make too much impact on wind speed.
 - b) Pearson correlation coefficient for **humidity & wind speed** was "-0.0111" with P-Value of "0.213". This means there is negative relation between humidity and wind speed but this relationship is not that solid.
 - c) Pearson correlation coefficient for **air pressure & wind speed** was "-0.023036" with P-Value of "0.00977". This means there is negative relation between air pressure and wind speed. And P-value is also making sure that there is strong evidence of air pressure negative impact on wind speed. And these results also convey that a decrease in air pressure will increase wind speed.
 - d) Pearson correlation coefficient for **wind direction & wind speed** was "0.0297" with P-Value of "0.0008". This means there is positive relation between wind direction and wind speed.

DS 670 – Assignment 13 - Final Manuscript

- **Discussions:**

In this paper, short-range forecasting average daily wind speed using inductive method was proposed. The principle of high-impact weather events substantiates the different places' interaction by atmosphere, hydrosphere, landmass, biosphere, etc. The main assumptions are:

1. Forecasting model (1) describes the wind speed time series adequately; coefficients k_1 , k_2 , and k_3 use to have the same sign as appropriate PPMCCs.
2. Inductive criterion (2) includes two polar parts – minimum of regularity plus maximum of conjunctions; hence, two training sequences' usage (classical self-organizing approach) is not necessary.

- **Conclusion:**

After the use of short-range forecast method there is significant improvement in the error rate of prediction. And consideration of relationship between different independent variables gave a good platform for prediction of **wind speed**. Statistical model used in this analysis is still under process of improvement. Soon, there will be inclusion of new parameters in the study that will be combination of existing parameters.