

Machine Learning Technique For Prediction Of Wind And Rainfall Using Underwater Measurement

Parimala¹

¹Assistant Professor, Department of Bachelor of Computer Application(Women), Sharnbasva University
Kalaburagi, Kalaburagi, Karnataka.
anupamababshetty@gmail.com

ABSTRACT

Analyzing underwater ambient acoustics may be used to measure precipitation as well as wind over ocean. For purposes of this research, machine learning approaches used to measure wind speed or precipitation rate were compared to empirical methods employed in prior studies using ambient ocean sounds collected by a PAL in Mediterranean. Genoa, Italy, was hit hard by two storms in June 2011 and May 2012, which resulted in significant coastal flooding. High-quality in situ data measured by a spar buoy on surface above PAL were used to train and validate model. There is a correlation coefficient of 0.95 between acoustic information with wind speed using machine learning models, which is a significant improvement over prior technique. Measurements of total precipitation for 12 occurrences were made using CatBoost plus random forest models, exhibiting mean and median errors of 28 and 34 percent and 18 and 17 percent, respectively, for both models. Using machine learning to analyze information from underwater acoustic recorders may assist enhance in situ measurements throughout seas worldwide.

Keywords- Wind, Rainfall, underwater measurements, passive aquatic listener (PAL), CatBoost

I. INTRODUCTION

New attention is being placed on ocean ambient noise monitoring in order to better understand how human activities affect marine animal ecology.

During World War II, preliminary investigations on undersea ambient noise were conducted that covered a broad variety of circumstances and locations of oceans. Ships' horns and sounds of breaking waves as well as marine creatures add to ambient noise, according to researchers. There is a low-frequency noise generated by ships travelling across the ocean, and Wenz [1] curves served as foundation for various prediction systems.

According to a 2003 National Research Council study[2], "noise linked with background din arising from a multiplicity of unexplained sources" was the new term of ambient noise.

According to Descriptor 11 of Marine Framework Strategy Directive (MFSD), "Introduction of energy," including underwater noise, must be monitored such that it "is at levels that do not negatively impact marine environment." [3] There have been specific regulations issued by European Commission since MFSD was issued, but for some descriptors, mature and technical working groups have been established more to develop descriptors as well as provide necessary regulations for implementation. For marine trash, they were published in 2011; for underwater noise, in 2012.

Both power as well as concentration of sound sources and propagation of sound from sources to receivers, which is controlled by environmental variables, play a role in underwater ambient noise measurements made by hydrophones (i.e., oceanographic dynamics, sound velocity profile, bathymetry). Ambient noise may be affected by a wide range of factors, including depth, time, location, and other variables, which necessitates experimentation to enhance quantification of sound sources based on ambient noise data.

II. LITERATURE SURVEY:

Y. L. Serra(2018): It is crucial to global climate system and one of essential climate variables (ECVs) in implementation plan for Global Climate Observing System (GCOS). 75 percent of all rainfall falls over open ocean, where there are few if any in situ measurements. For an ECV to be met, validation of existing satellite methods for rainfall estimation using in situ measurements is still necessary, despite significant improvements over last two decades. Here we take a look at present state of Tropical Moored Array (TMA) open-ocean rainfall observations throughout tropical Pacific as well as look forward to difficulties and possibilities of TPOS 2020 Project, which is now under way. Over last 15 years, TMA has collected rainfall data, although this data has been substantially decreased in recent years. This is an ideal opportunity to revisit TMA rainfall data and their

role in redesign of TMA as a "full flux" platform, which includes measurements of radiation, rainfall, momentum, latent, as well as sensible heat fluxes.

B. B. Ma and J. A. Nystuen(2005) : As a climatic characteristic for oceanic and atmospheric research, rainfall over ocean is an essential one to keep an eye on. Using underwater sound as a substitute for traditional deposition rain gauges has been created. Precision of sound pressure level (SPL) measurement is critical to passive monitoring of ocean rainfall using ambient sound. As a result, absolute correction of hydrophone is desired, but difficult to obtain due to fact that measurement geometry over ocean often does not meet laboratory calibration procedure geometry. In other words, if one considers that sound signal that comes from wind is universal, then wind signal may be utilised to calibrate an instrument. Tropical Atmosphere Ocean (TAO) project has been collecting ambient sound spectra for over 90 buoy months since 1998. It is possible to get instrument noises and the sensitivity prejudice of each ARG by using wind speed method developed by Vagle et al. Pure geophysical signals are recovered using an auditory discrimination method. Using data from ARG and a self-siphoning rain gauge positioned on same mooring, a novel single-frequency rainfall-rate algorithm is suggested. Other rain gauges from R.M. Young and Tropical Rain Measuring Mission (TRMM) product 3B42 will be used to evaluate auditory discrimination method as well as rainfall algorithm. Long-term (years) and short-term (hours) comparisons of acoustic rainfall accumulation indicate same outcomes in both cases.

S. Pensieri, R. Bozzano, J. A. Nystuen, E. N. Anagnostou, M. N. Anagnostou and R. Bechini(2015) : Wind and rain events over ocean may be examined using oceanic ambient noise data to cover current gap in accurate weather stations at sea. By using long-term synergistic observations from an acoustic recorder and surface sensors (such as rain gauges and anemometers), the Ligurian Sea Acoustic Experiment was able to assist the evaluation of rainfall rate as well as wind speed quantification approaches used in previous research. There were a number of high and low-wind and rainfall occurrences, as well as two storms that flooded the area around La Spezia and Genoa in 2011 during research period. It is now easier to identify and forecast wind and rain at sea because to availability of high-resolution meteorological data in situ. Data shows an excellent correlation between rainfall and wind estimations supplied by algorithm as well as in situ observations, demonstrating that readings given by passive acoustic equipment may be used to offer early warning for coming coastal storms, which is one of leading causes of flooding in Mediterranean coastal regions.

S. Pensieri, R. Bozzano, M. N. Anagnostou, E. N. Anagnostou, R. Bechini and J. A. Nystuen(2013): Underwater noise plays an important role in study of acoustics in marine environments and acoustic oceanography since it is an intrinsic aspect of marine environment as well as a vital parameter for enhancing sustainability and maintaining marine ecosystem. Even more importantly, long-term monitoring of wind and rain events by analysing underwater sound measurements could greatly enhance our understanding of these atmospheric processes in open-ocean areas where climatic factors confine our ability to obtain in-situ observations via surface buoys. W1-M3A spar buoy, equipped with a Passive Aquatic Listener (PAL) sensor, was put in Ligurian Sea to measure natural sound sources such as wind, rain, and ship crossings. Findings of this study are presented in this report. Meteorological or ship traffic data from Ligurian basin are directly compared to PAL data to ensure that estimations are accurate to within a few percent.

J. Yang, U. of Washington, S. Riser, J. Nystuen, W. Asher and A. Jessup(2015): Knowledge global hydrological cycle requires an understanding of rainfall intensity as well as spatial-temporal distribution across ocean. On other hand, because to station motion as well as flow distortion and rainfall unpredictability, it has been difficult to detect rain over oceans accurately. In order to detect and measure rainfall, underwater acoustical rain gauges use loud and unique underwater sound made by raindrops on ocean surface. Presented here are details on how an apparatus that employs underwater external noise to quantify rainfall frequency as well as wind speed works, as well as some of the findings. US National Aeronautics and Space Administration-sponsored Salinity Factors in an Upper-ocean Regional Study (SPURS) field experiment in North Atlantic Ocean used PAL equipment installed on a buoy at Ocean Station P and 13 Argo profilers. Over SPURS research area specified by Argo profilers, PALs offer near-continuous observations of rain rate and wind speed. There is high agreement in rain rate and wind speed between PAL data and other methods of rain and wind measurement, particularly direct in situ observations as well as satellite measurements.

G. Ireland, M. Volpi and G. Petropoulos (2015): SVMs and regularised kernel Fisher's discriminant analysis (rkFDA) machine learning supervised classifiers were used in this work to recover flooded areas from optical Landsat TM images. To use a case study of a riverine flood occurrence in a diverse Mediterranean area in 2010 for which TM imagery was available, both approaches were assessed. Both linear and non-linear (kernel)

versions of two classifiers were used in their development. Classifiers' capacity to map the size of a flooded region was evaluated using criteria for classification accuracy. According to results, rkFDA outperformed SVMs when it came to accurately finding flooded pixels and missed detections. SVMs, on other hand, revealed less false alarms due to flooding. Classification accuracy was higher using non-linear rkFDA classification approach (OA = 96.23%, K = 0.877) than with linear method. Both strategies exceeded typical NDWI thresholding (OA = 94.63, K = 0.818) by around 0.06 K points. There was only a slight improvement in overall accuracy between rkFDA and SVMs classifications, but both classifiers significantly outperformed thresholding algorithm in other accuracy measures (for example, producer accuracy for the "not flooded" class was 10.5% less accurate for thresholding algorithm than for classifiers, and average This paper shows that supervised machine learning may be used successfully to categorise flooded regions in Landsat images, something that has received very little research to date. Landsat data is free and provides worldwide coverage, therefore findings of this research may be used to explore potential of using such data in an economically feasible manner to record additional large flood occurrences from space.

D. A. Sachindra, K. Ahmed, M. M. Rashid, S. Shahid and B. J. C. Perera(2018): In order to downscale reanalysis data to monthly precipitation at 48 observation sites spread over Australian state of Victoria, belonging to wet, intermediate, and dry climatic regimes, statistical models were built. Calibration and validation of models for monthly downscaling for each station were done using four machine learning techniques, including genetic programming (GP), artificial neural networks (ANNs), support vector machines (SVM), and relevance vector machines, for the years 1950–1991 and 1992–2014, respectively (RVM). It was discovered that downscaling models, regardless of climatic regime or machine learning approach, tend to better replicate average and underestimate standard deviation as well as maximum of observed precipitation (relative to other statistics). Downscaling models, regardless of climatic regime and machine learning approach, exhibit an overestimation of low to mid percentiles of precipitation and an underestimation of high percentiles of precipitation at majority of stations studied (i.e. above the 90th percentile). Drier climates were more likely to show an overestimation tendency for precipitation in the low to mid percentiles, regardless of machine learning approach. Using RVM or ANN over SVM or GP for creating downscaling models for a subject such as flood prediction is advised based on findings of this experiment. GP, ANN, and SVM may also be used in constructing downscaling models for a research such as drought analysis, but RVM is preferable since it considers the low extremes of precipitation while developing models. In addition, it was discovered that Polynomial kernel provided greatest performance for precipitation downscaling models based on SVM and RVM, regardless of climatic regime.

A. Belayneh, J. Adamowski, B. Khalil and J. Quilty(2016): Drought conditions in Ethiopia's Awash River Basin were predicted using a combination of machine learning models and group methodologies. When used in conjunction with bootstrap and boosting ensemble approaches, wavelet transformations may help create highly accurate artificial neural network (ANN) as well as support vector regression model predictions for drought. As a pre-processing approach, use of wavelet analysis was shown to enhance drought forecasts. Short-term and long-term drought conditions are represented by Standardized Precipitation Index, which was predicted using aforementioned models, and these SPI values. In order to compare performance of all models, RMSE, MAE, and R2 were used. Results showed that boosting ensemble approach consistently increased correlation among observed as well as projected SPI. To further enhance model predictions, we used wavelet analysis. More accurate predictions were obtained using wavelet boosting neural network (WBS-ANN) and wavelet boosting stochastic vector autoregressive network (WBS-SVR).

S. Pensieri, R. Bozzano and M. E. Schiano(2010): In Mediterranean Sea, wind above water's surface plays a vital role in a wide range of research areas (i.e., climatological, meteorological and oceanographic studies). An anemometer deployed on an offshore ODAS Italia 1 buoy and wind vectors received from SeaWinds scatterometer on board NASA QuikSCAT satellite are compared in this study. Central Ligurian Sea, a region with orographic restrictions and temperature contrasts between land and sea, generates a very changeable wind field. This platform is anchored in Central Ligurian Sea. From July 2006 through June 2007, work is included.

QuikSCAT's wind vectors meet accuracy standards for high wind speeds, however direction assessment is less accurate at lower wind speeds, according to comparison.

These findings are in line with earlier research, but examination uncovered certain difficulties that had gone unaddressed. Most important issue is lack of information. Only twice a day does satellite travel over Ligurian Sea, because data collected in wet or windy conditions is invalid (wind speed less than 3 ms⁻¹). To observe wind

fields but also their development across this area, temporal sample may not be sufficient, since calm conditions are common and severe disturbances are often accompanied with rain. When genuine wind speed is less than 3 ms⁻¹, comparison indicates that it is impossible to distinguish bogus estimates derived from QuikSCAT without a reference at sea.

III. OBJECTIVES:

1. To implement best approach available for detection and prediction of rainfall in certain area of country supervised machine learning concept.
2. To let users, know if each year data is analysed accurately and efficiently
3. The accuracy of the prediction increasing
4. Acoustic underwater measurement must be considered in the prediction.

IV. SYSTEM DESIGNS:

System Architecture

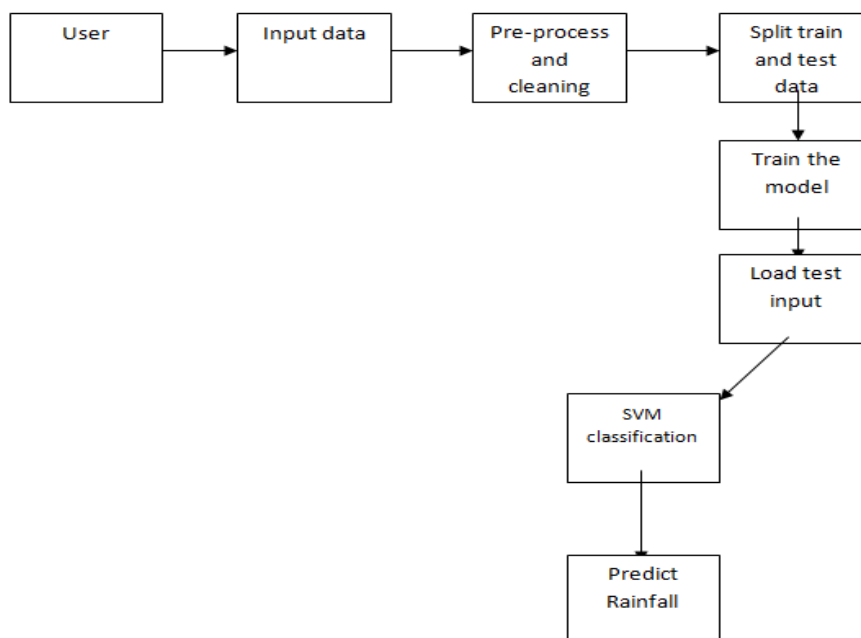


Fig: 1 System Architecture

DFD:

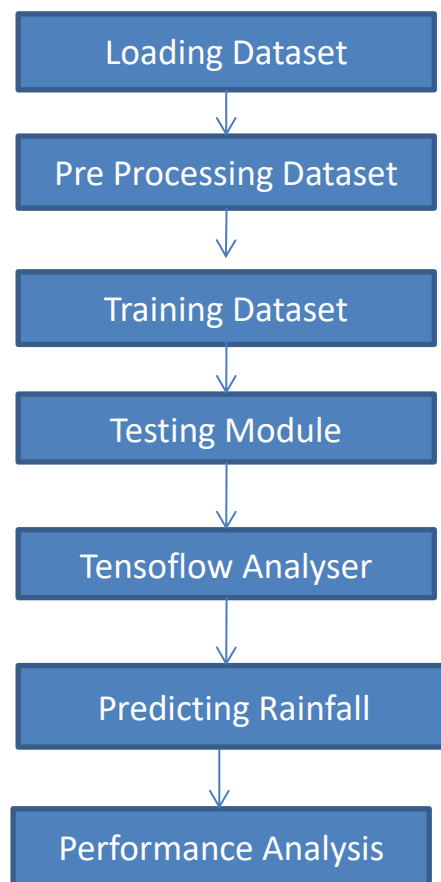


Fig: 2 Data Flow Chart:

Sequence Chart:

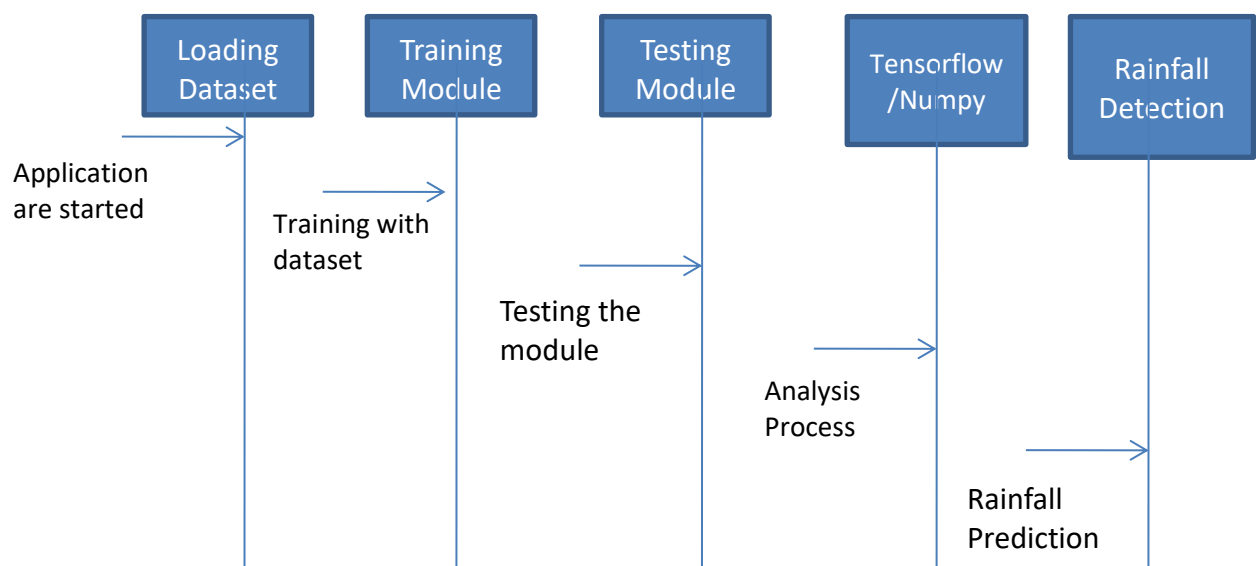


Fig: 3 Sequence Chart:

Use case Diagram:

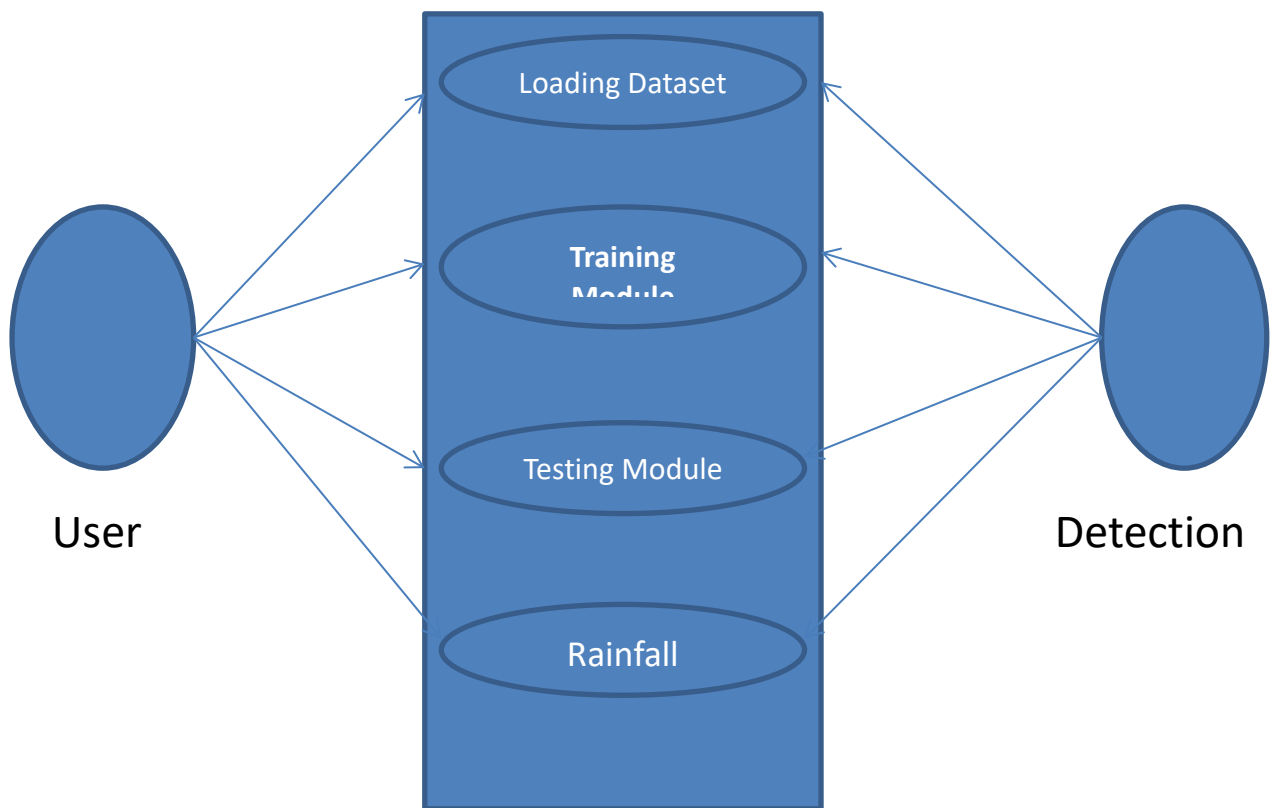
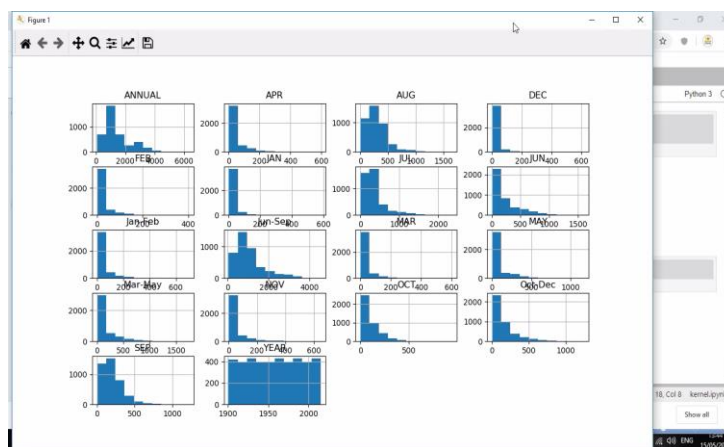
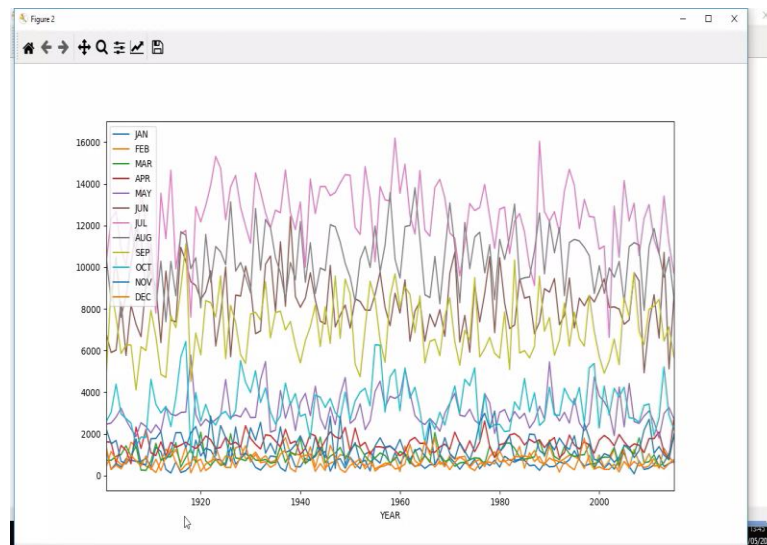


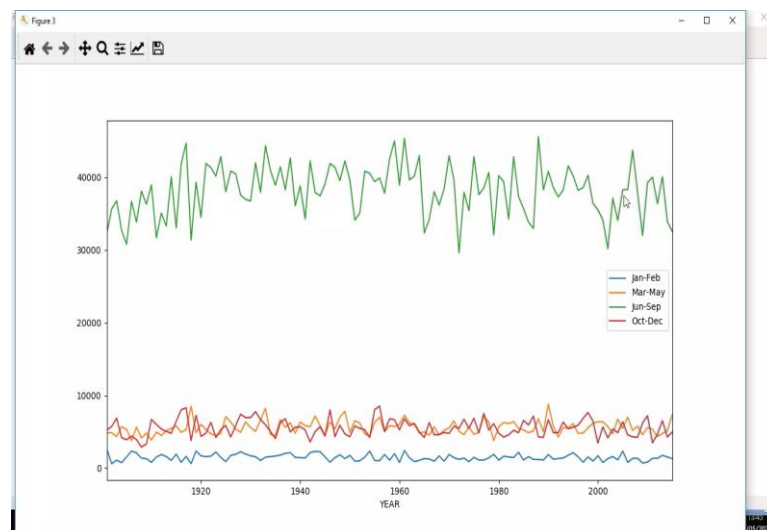
Fig:4 Use Case Diagram:



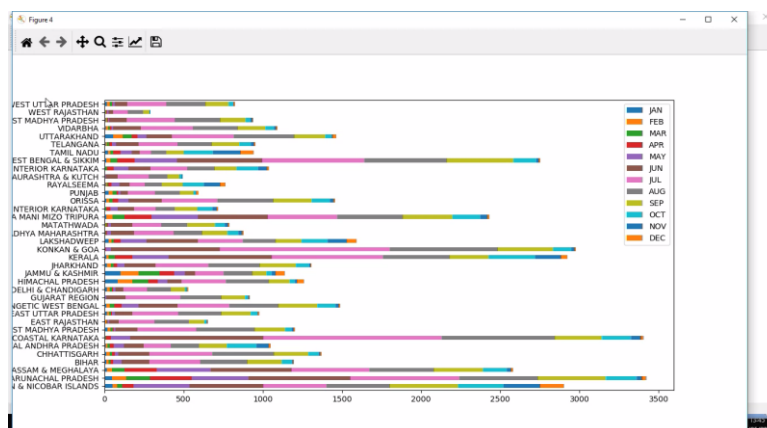
Screen 1: Histogram Chart for the Rainfall Details.



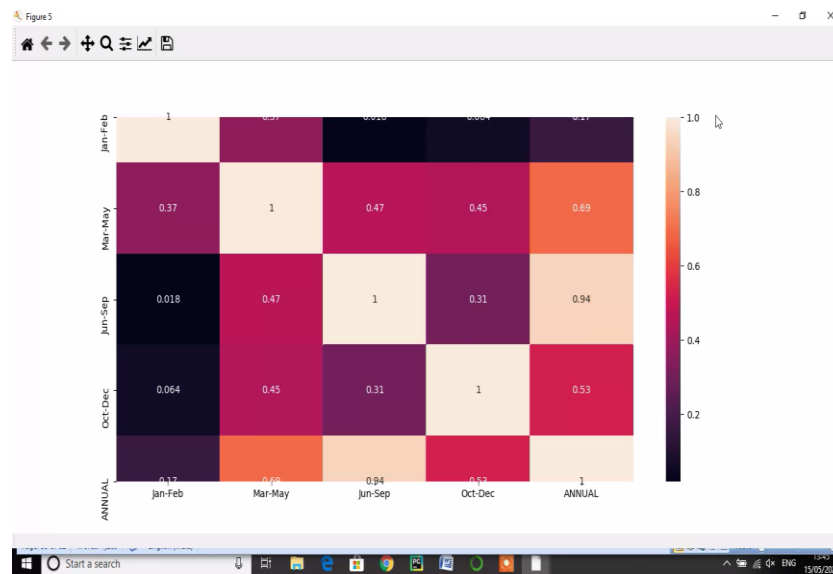
Screen 2: Line chart Details



Screen 3: Line chart foe 1920 to 2000.



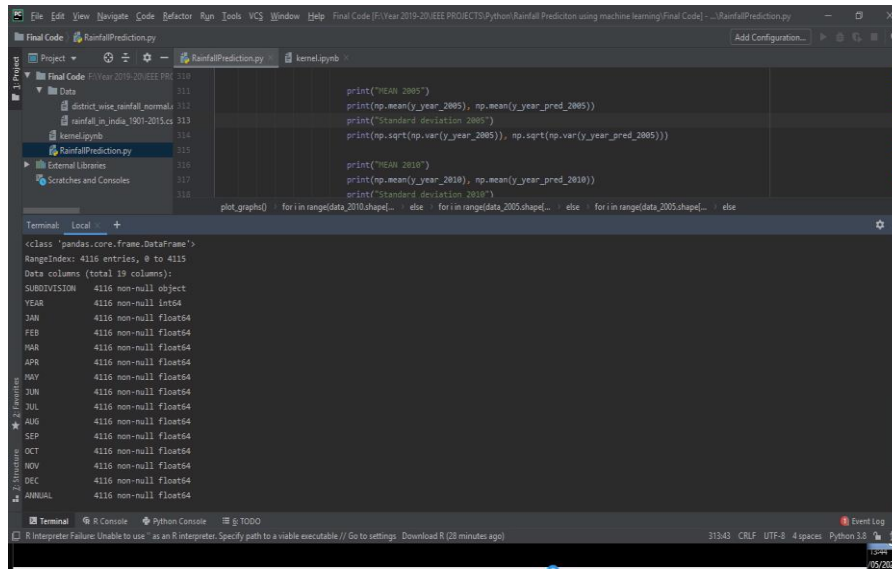
Screen 4: bar chart details of the states where how much of rainfall happened.



Screen 5: matrix diagram for the rainfall analysis.



Screen 6: Heat-map diagram information



```

File Edit View Navigate Code Refactor Run Tools VCS Window Help Final Code [F:\Year 2019-20\IEEE PROJECTS\Python\Rainfall Prediction using machine learning\Final Code] - ...RainfallPrediction.py
Project Final Code RainfallPrediction.py kernel.py
Data
    district wise rainfall normal 312
    rainfall_in_india_1901-2015.csv 313
    kernel.py 314
    RainfallPrediction.py 315
External Libraries
Scratches and Consoles
Terminal Local +
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4116 entries, 0 to 4115
Data columns (total 19 columns):
SUBDIVISION    4116 non-null object
YEAR           4116 non-null int64
JAN            4116 non-null float64
FEB            4116 non-null float64
MAR            4116 non-null float64
APR            4116 non-null float64
MAY            4116 non-null float64
JUN            4116 non-null float64
JUL            4116 non-null float64
AUG            4116 non-null float64
SEP            4116 non-null float64
OCT            4116 non-null float64
NOV            4116 non-null float64
DEC            4116 non-null float64
ANNUAL         4116 non-null float64

```

Screen 7: Command line output for the rainfall prediction analysis.

V. TESTING METHODOLOGIES

Listed below are many testing techniques:

- Unit Testing.
- Integration Testing.
- User Acceptance Testing.
- Output Testing.
- Validation Testing.

Unit Testing

Module is smallest unit of software design which can be tested in unit testing. For optimum error identification and coverage, unit testing exercises individual control routes in a module's control hierarchy. As a whole, this test ensures that each module works as intended. Thus, name "Unit Testing" was coined.

All of modules are tested separately, and their interfaces are checked to make sure they match design specifications. There are predicted outcomes for every significant processing route. Error-handling procedures have been thoroughly tested, as well.

Integration Testing

Testing for both verification and programme creation is part of integration testing. An extensive set of tests are carried out when programme is incorporated. Unit tests are used to develop a programme structure that is determined by architecture of application.

Integration testing may be broken down into following categories:

1. Top Down Integration

Building a program's structure in this way is a gradual process. Starting with main programme module, modules are integrated by descending via control hierarchy one module at a time. Subordinate modules to primary programme module are integrated into framework either in depth or breadth initially.

By starting with main module and working down, this approach allows product to be thoroughly tested without having to modify any stub code along way.

2. Bottom-up Integration

Modules at lowest level of programme structure are used to begin building as well as testing. Processing for subordinate modules is always accessible since modules are integrated from bottom to top. Stubs are no longer necessary. With following phases, bottom-up integration approach may be put into action:

- Some software sub-functions are performed by groups of low-level modules integrated into groups.
- Test case input, output are coordinated by writing a driver (i.e., a control programme for testing).
- There is a test of cluster.
- Moving up programme hierarchy, drivers are deleted and clusters are consolidated.

Using a bottom-up method, each module is tested on its own before being incorporated into a larger module being evaluated for its functionality.

User Acceptance Testing

Key to any system's success is its acceptance by its users. User acceptability is assessed by regularly communicating with and making adjustments to prospective system users while system is still being developed. There is a nice user interface that is easy to understand even for those who have never used system before.

Output Testing

A system cannot be of any use if it does not provide appropriate output in stated format, hence next step is to do output testing when validation testing is complete. Outputs created or shown by system in review may be tested by asking users what format they prefer. Output format is thus seen from two perspectives: on-screen and printed.

Validation Checking

Following fields are subjected to verification checks.

Text Field:

Only characters less than or equal to the field's size may be entered into text field. Some tables include alphanumeric text fields, whereas others have alphabetic text fields. An error message is shown whenever an incorrect entry is made.

Numeric Field:

Numerical values in this field must fall within the range of 0 to 9. Error warnings are shown whenever a character is entered incorrectly. Accuracy as well as performance requirements are verified for each module. Sample data is used to run each module through its paces. A single system is constructed from components that have been thoroughly evaluated on their own. Program defects can only be discovered by analysing how program's output compares to its input. Testing must be designed in such a way that each of criteria is tested independently.

A successful test is one which reveals flaws in system by showing deficiencies in the incorrect data.

Preparation of Test Data

Take several types of test data to test aforementioned. When a system is being tested, it is essential that test data be prepared. It is only when test data is prepared that system under investigation is tested. Errors discovered and repaired during testing by employing test data are documented for future use, as are corrective actions taken.

Using Live Test Data:

Data that is really taken from an organization's files is known as live test data. It's not uncommon for programmers or analysts to ask end users to provide a collection of data from their daily routines when a system is largely built. Systems expert then utilises this data to test system in part. Programmers and analysts may also pull real time data from files but also insert it manually.

It is challenging to get enough live data to do comprehensive testing. As a result, even if actual data submitted are in fact typical, the system's performance will not be tested for all different pairs or formats that

may be entered into it. As a result, a real system test is hampered by this tendency toward normal values, which misses the failure modes that are most likely to lead to system failure.

Using Artificial Test Data:

Because they may be constructed to test all possible forms and values, artificial test cases are used just for testing reasons. To put it another way, fake data, which can be swiftly created by an information systems department data generating utility application, allows testing of all login as well as control channels via programme.

Most successful test programmes incorporate data supplied by people other than programmers. A team of independent testers may devise a testing strategy based on system requirements.

Acceptance of "Virtual Private Network" package was based on its compliance with the software requirements specification.

USER TRAINING

For each new system, user training is necessary to ensure that it can be effectively used by individuals whose system was originally intended. Potential users were shown how project normally functioned for this purpose. Using this system is a simple since it's designed for folks who already have a decent grasp of computers.

MAINTAINENCE

Correcting code and design flaws are only two examples of this kind of work. During system development process, we have more precisely identified user's needs in order to avoid need for long-term maintenance. This system has been designed to meet demands of greatest degree feasible based on specifications. In future, as technology advances, it may be feasible to include a wide range of additional functions. In order to make maintenance simpler, code and design is basic and straightforward.

TESTING STRATEGY:

A strategy for system testing combines testing phase cases and design methodologies into a well-planned set of phases that lead to effective development of software. Coordination between test planning, test design, execution and data gathering and assessment is essential. Low-level tests that verify that a tiny portion of source code has been implemented successfully as well as high-level tests that check significant system functionalities versus user requirements must be included in a testing strategy.

Assurance of software quality is vital, and testing is last step in design and development process. Software's testing is an intriguing quirk. Thus, proposed system is subjected to a series of tests before it is suitable for user testing.

SYSTEM TESTING:

It is essential that software be integrated with other system components after validation. Testing ensures that all components are functioning properly so system capacity is met. It also looks for conflicts between system's current specs as well as documentation and thus its original aim.

UNIT TESTING:

To ensure that modules meet their design standards, they are put through a series of tests in form of unit testing. Purpose of unit testing is to verify code that was written during coding process, and therefore to test module's internal logic. When testing Conrail pathways, defects inside modules' boundaries are discovered by following comprehensive design description. During programming phase, this testing is undertaken out. Accordingly, each module was determined to be performing as predicted in this sort of testing process.

Consideration will be given to most recent technological breakthroughs in future. Many networking components will be made generic as part of technological setup so that they may be used or interacted with by future applications. This project has a lot to look forward to in years to come.

VI. CONCLUSION AND FUTURE WORK:

Using a combination of machine learning with forecasting approaches, proposed study aims to predict rainfall. Even though there are numerous variables that influence rainfall, we are able to achieve outstanding classification accuracy with only a few. In addition, we discovered that even after categorising rainfall into eight distinct groups, we still got respectable accuracy. This method is used to verify projected parameters. As shown by test results, ARIMA performs best for highest temperature, Neural Network performs best for lowest temperature, SVR performs best for humidity levels but also wind speed. A classification's validity may be evaluated by looking at metrics like accuracy, precision, and recall. Random forest is best classifier for rainfall classification, according to ROC curve for all classifiers.

It is necessary to explore how different meteorological characteristics impact Rainfall forecast since rainfall is reliant on several factors. Using hourly data and a variety of characteristics, we can do same thing to predict how much rain will fall in following hour. This kind of model may be built on a big data framework to allow for quicker computation while maintaining greater accuracy as part of a research.

REFERENCES

1. Mithila Sompura Aakash Parmar, Kinjal Mistree. Machine learning techniques for rainfall prediction: A review. International Conference on Innovations in information Embedded and Communication Systems, 2017.
2. Nishchala C Barde and Mrunalinee Patole. Classification and forecasting of weather using ann, k-nn and naïve bayes algorithms.
3. Debasish Basak, Srimanta Pal, and Dipak Chandra Patranabis. Support vector regression. Neural Information Processing-Letters and Reviews, 11(10):203{224, 2007.
4. Leo Breiman. Random forests. Machine learning, 45(1):5{32, 2001.
5. KK Chowdari, R Girisha, and KC Gouda. A study of rainfall over india using data mining. In Emerging Research in Electronics, Computer Science and Technology (ICERECT), 2015 International Conference on, pages 44{47. IEEE, 2015.
6. Pinky Saikia Dutta and Hitesh Tahbilder. Prediction of rainfall using data mining technique over assam. IJCSE, 5(2):85{90, 2014.
7. G Gregoire. Multiple linear regression. European Astronomical Society Publications Series, 66:45{72, 2014.
8. Mina Mahbub Hossain and Sayedul Anam. Identifying the dependency pattern of daily rainfall of dhaka station in bangladesh using markov chain and logistic regression model. 2012.
9. Rob J Hyndman. Moving averages. In International Encyclopedia of Statistical Science, pages 866{869. Springer, 2011.