**Title: Assessing Water Quality And Management**

**CONTENT**

|  |  |  |
| --- | --- | --- |
| **S No.** | **Topic** | **Page No.** |
| 1. | Abstract | 2 |
| 2. | Introduction | 3 |
| 3. | Architecture Diagram with explanation | 9 |
| 4. | UML class diagram with explanation | 11 |
| 5. | DB design in graphical form |  |
| 6. | Tools Used |  |
| 7. | Test Cases |  |
| 8. | Conclusion |  |
| 9 | References |  |

**ABSTRACT**

Human beings are dependent on the natural resources that stands for its quality. Climatic changes and environmental impacts have always been under observation. The quality of the products is always measured before use. Water, an inevitable resource, has got a serious significance in checking its quality due to the influence of various external factors like industrial effluents, acid rain etc.

The nature of drinking water has consistently been an incredible ecological determinant of wellbeing concern around the world. A protected and safe stock of drinking water is principal to general wellbeing. Water quality is generally ascertained by the number of contaminants in it. Accurate detection of contaminants is often required for planning and implementing mitigation measures to protect water supplies. Groundwater is arising as a basic issue for urban areas and towns around the planet. It is estimated that approximately people who use groundwater for drinking are one-third of the world’s population.

Nowadays Water Quality management has become a major problem of overall water resources management. The people who are living in rural areas will depend on low levels of water quality. That poor people don’t know what the chemicals contained in that low-level water. More pollutants will be generated if the populations and economics grow and finally, they end up with the surface and groundwater bodies. Every living thing or organism requires water not for just sufficient quantities but also contains quality to survive healthy.

Ever increasing population, urbanization and modernization are posing problems of sewage disposal and contamination of surface waters like lakes. Natural water gets contaminated due to weathering of rocks, leaching of soils and mining processing, etc. Various types of problems in lake which cause nutrient enrichment in lake have been reviewed. Land use change and longer growing seasons could increase the use of fertilizers with subsequent leaching to watercourses, rivers and lakes, increasing the risk of eutrophication and loss of biodiversity. Water quality can be assessed by various parameters such as BOD, temperature, electrical conductivity, nitrate, phosphorus, potassium, dissolved oxygen, etc. Heavy metals such as Pb, Cr, Fe, Hg, etc. are of special concern because they produce water or chronic poisoning in aquatic animals. Harmful algal blooms are becoming increasingly common in freshwater ecosystems globally. Pollution by plastic debris is an increasing environmental concern in water bodies, where it affects open-water, shoreline and benthic environments. Surface water densities of plastics are as high as those reported for areas of litter accumulation within oceanic gyres. It is recommended that pollution prevention and water re-use should be adopted in combination with the recycling of nutrients in controlled urban agriculture.

The main aim of this project is to identify the values, which are more affected by scanning the dataset and taking care to reduce contaminants in that rural area. Also, this helps people register complaints, know the status of water in their area.

**INTRODUCTION**

Nowadays Water Quality management has become a major problem of overall water resources management. The people who are living in rural areas will depend on low levels of water quality. That poor people don’t know what the chemicals contained in that low-level water. More pollutants will be generated if the populations and economics grow and finally, they end up with the surface and groundwater bodies. Every living thing or organism requires water not for just sufficient quantities but also contains quality to survive healthy. The main aim of this project is to identify the values, which are more affected by scanning the dataset and taking care to reduce contaminants in that rural area. Drinking water quality can be determined by performing water quality analysis, which is a major component of the water distribution system.

● Water quality analysis is considered as the most important component during drinking water quality analysis as it has a heavy impact on public health and industrial use.

● The causes of water quality reduction are multi-folded and some of them include the following:

1. Poor sanitation.

2. Insufficient resources for monitoring and maintaining water quality management systems.

3. Salt water intrusion.

4. Chemical contamination.

Water quality assessment is the overall process of evaluation of the physical, chemical and biological nature of the water, whereas water quality monitoring is the collection of the relevant information. This Project concentrates on the whole assessment process, in different types of water bodies (complaints received). The details of monitoring methods and approaches which can be applied in the field

Here we are designing a portal in a way that we can interact with the people around us through this complaint module, so we did some background work on web development. Through the website we will be able get some data about the complaints so therefore we have learnt about data acquisition. In order to classify the data, we need to do data cleaning, data transformation, data reduction, data pre-processing parameters so we learnt some data mining techniques to do the mentioned above.

**Chosen Model Type:**

The model that will be suitable for Assessing Water Quality and Management is Water Fall model.

The reason why we propose spiral model is that Assessing Water Quality and Management software is a large sophisticated community that demands high flexibility and response time. It includes estimating the cost, schedule and report generation for the iteration.

Cost estimation becomes easy as the prototype building is done in small fragments. Since, it is a large project the use of water fall model is highly needed.

Reasons for choosing the Water fall model is:

### 1. It Uses clear structure

When compared with other methodologies, Waterfall focuses most on a clear, defined set of steps. Its structure is simple—each project goes through these steps:

* Requirement gathering and documentation
* System design
* Implementation
* Testing
* Delivery/deployment
* Maintenance

Teams must complete an entire step before moving onto the next one, so if there are roadblocks to completion, they’re brought to light right away. Half-finished projects are less likely to get pushed aside, leaving teams with a more complete, polished project in the end.

In addition to being clear, the progression of Waterfall is intuitive. Unlike Six Sigma or Scrum, Waterfall does not require certifications or specific training for project managers or employees. If you visually outline the process at the beginning using Lucid chart and explain the methodology, team members will be able to jump into the Waterfall system without a steep learning curve slowing their progress.

### 2. Determines the end goal early

One of the defining steps of Waterfall is committing to an end product, goal, or deliverable at the beginning, and teams should avoid deviating from that commitment. For small projects where goals are clear, this step makes your team aware of the overall goal from the beginning, with less potential for getting lost in the details as the project moves forward.

Unlike Scrum, which divides projects up into individual sprints, Waterfall keeps the focus on the end goal at all times. If your team has a concrete goal with a clear end date, Waterfall will eliminate the risk of getting bogged down as you work toward that goal.

### 3. Transfers information well

Waterfall’s approach is highly methodical, so it should come as no surprise that the methodology emphasizes a clean transfer of information at each step. When applied in a software setting, every new step involves a new group of people, and though that might not be the case at your company, you still should aim to document information throughout a project’s lifecycle. Whether you’re passing projects off at each step or experience unexpected personnel changes, Waterfall prioritizes accessible information so new additions to the team can get up to speed quickly if needed.We can maximize the system benefits from this characteristic of Waterfall by staying organized with the right process.

**Modules:**

1. **Complaint Module:**

In the complaint module, the complaints are being handled and analyzed by officials once they get registered by the localities.

● Here we are designing a portal in a way that we can interact with the people around us through this complaint module.

People can raise their respected queries or put any complaints with respect to quality of water or any problems related to their water and mention their panchayat details.

● So the complaints are taken into consideration and actions are taken/provided by the officials accordingly.

1. **Data Processing:**

**i. Data Acquisition:**

Data acquisition is often referred to as the process of digitizing data from the world around us so it can be displayed, analyzed and stored in a computer.

⮚ So, in data acquisition we will collect data

⮚ When a citizen wants to raise a complaint that can be done through the web portal.

⮚ In the web portal we will be able to see the raised complaints by the people about the quality terms or any other regarding water issues.

⮚ This information (data) is then collected, analyzed and then stored in the computer.

⮚ During this process the ontology is used to navigate the user while collecting information using data type properties.

⮚ And also, custom labels are used to address the user with specific information if another change takes place before the current change is acquired.

⮚ The parameters that should be filled in the complaint form by people are state name, district name, block name, panchayat, village name, habitation, year and quality parameter.So, using a data acquisition system allows us to obtain valuable information of the reality to improve the performance of the company and to increase the economic benefit. It is a must done process as it provides greater control over an organization's process and faster response to failures that may occur.

**ii. Data Preprocessing:**

Steps Involved in Data Preprocessing:

1. **Data Cleaning**: The data can have many irrelevant and missing parts. To handle this, data cleaning is done. It involves handling missing data and noisy data.

a) Missing data: It arises when some data is missing in the data. It can be handled by ignoring the tuples or fill the missing values by attribute mean or the most probable value.

b) Noisy data: It is meaningless data that cannot be interpreted by machines. It can be handled by various methods like binning, regression and clustering.

2. **Data Transformation:** This step is taken in order to transform the data in appropriate forms suitable for the project. This may involve Normalization, Attribute Selection, Discretization and Concept Hierarchy Generation.

3. Data Reduction: It increases the storage efficiency and reduces data storage and analysis costs. There are various steps involved in data reduction. They are:

a. Data Cube Aggregation.

b. Attribute Subset Selection.

c. Numerosity Reduction.

d. Dimensionality Reduction.

**iii. Data Processing Parameters:**

Data has many parameters to consider, some data have to be filtered based on parameters such as locality, state name, district name, panchayat name, village name, quality parameters such as minerals and year. This data is processed based on the parameters. Process metrics can be used to increase the effectiveness of development and maintenance. This includes the effectiveness of removal during development, pattern of testing defect, and the response time of the fix process. Project metrics describes the various characteristics in a project and execution. This includes the total number of software developers, tester, cost, schedule, and productivity.

1. **Report:**

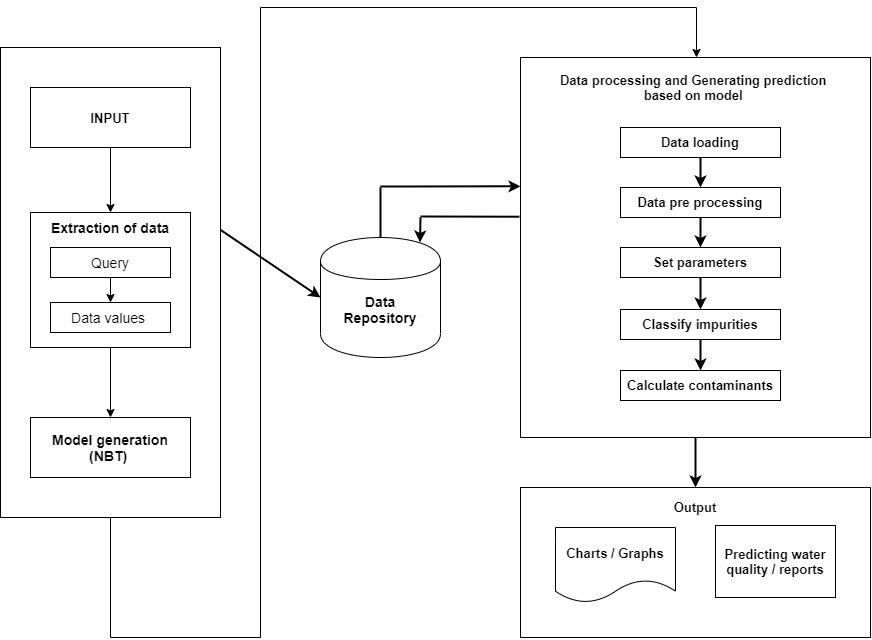
Reporting the final results obtained from Naive Bayes theorem, metric analysis and other graphs generated. This includes defining the report format, data extraction and reporting cycle, reporting mechanisms, distribution, and availability. The reporting cycle defines how often the report is generated and when it is due for distribution. For example, root cause analysis metrics may be triggered by some event, like the completion of a phase in the software development process. Other metrics like the defect arrival rate may be extracted and reported on a daily basis during system test and extracted on a monthly basis and reported quarterly after the product is released to the field.

A software development project is an experiment that is based on a number of assumptions, theories and estimates. To confirm these measurements are taken for all significant processes and then analyzed. The results of the analysis are used to make decisions.

1. **Analysis and Control:**

Analysis and Control manages and controls the overall assessing and predicting water quality process. This module identifies the work to be performed and develops the schedules and costs estimates for the effort. It coordinates with all activities and assures that all are operating from the same set of requirements, agreements and design iteration. It is the center for configuration management throughout the process. It examines the outputs of the other activities and conducts independent studies to determine which of the alternate approaches is best suited to the software. It finds outcomes when results of one activity require the action of another activity and directs the action to be performed. As the process progresses, trade-off studies and system effectiveness analyses are performed in support of evaluation and selection processes of the other activities. Risk identification studies are conducted to aid in Risk Management. Analyses also identify critical parameters and paths to be used in progress measurement. It specifies the parameters to be tracked for progress measurement. It conducts evaluation and reports progress.

**ARCHITECTURE DIAGRAM WITH EXPLANATION**



**Explanation of the Architecture:**

* **Input:** we are importing the data set first
* **Data repository:** A data repository is also known as a data library or data archive. This is a general term to refer to [a data set](https://www.informatica.com/services-and-training/glossary-of-terms/data-repository-definition.html#fbid=7ccCEj97R6Q) isolated to be mined for data reporting and analysis.The [data repository](http://www.infotoday.com/cilmag/apr16/Uzwyshyn--Research-Data-Repositories.shtml) is a large database infrastructure —  several databases — that collect, manage, and store data sets for data analysis, sharing and reporting.
* **Data extraction:** Data extraction is the process of collecting or retrieving disparate types of data from a variety of sources, many of which may be poorly organized or completely unstructured. Data extraction makes it possible to consolidate, process, and refine data so that it can be stored in a centralized location in order to be transformed. These locations may be on-site, cloud-based, or a hybrid of the two.
* **Model generation:** The Naïve Bayes classifier is probably the most straightforward way to deal with the classification task that is as yet equipped for giving sensible accuracy. Bayesian inference, of which the Naïve Bayes classifier is a particularly simple example, is based on the Bayes rule that relates conditional and marginal probabilities. Implementing such Naïve Bayes Algorithm to classify based on a decision attribute which is formed by mining the responded at a set from several participants and assigning a class is the approach that we have used following this procedure in this project.
* **Data processing and generating prediction based on model:**

1. **Data loading:** Loading the input data into the model
2. **Data Pre-processing:** Steps Involved in Data Pre-processing:

1. Data Cleaning: The data can have many irrelevant and missing parts. To handle this, data cleaning is done. It involves handling missing data and noisy data.

2. Data Transformation: This step is taken in order to transform the data in appropriate forms suitable for the project. This may involve Normalization, Attribute Selection, Discretization and Concept Hierarchy Generation.

3. Data Reduction: It increases the storage efficiency and reduces data storage and analysis costs. There are various steps involved in data reduction.

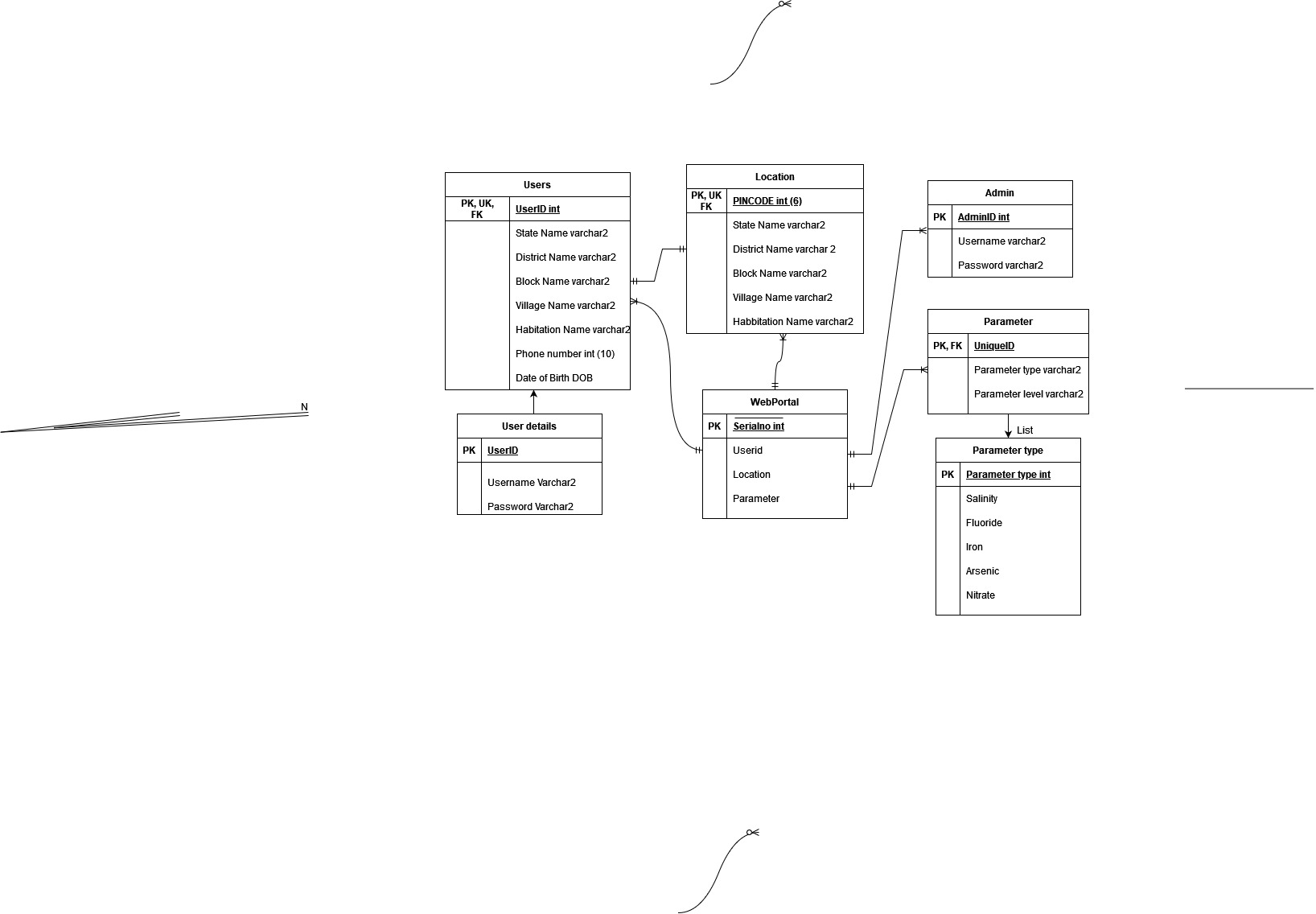
4. Data Processing Parameters: Data has many parameters to consider, some data have to be filtered based on parameters such as locality, state name, district name, panchayat name, village name, quality parameters such as minerals and year.

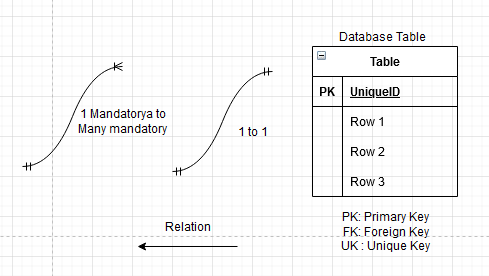
1. **Set parameters:** This data is processed based on the parameters. Process metrics can be used to increase the effectiveness of development and maintenance. This includes the effectiveness of removal during development, pattern of testing defect, and the response time of the fix process.
2. **Classify parameters:** Classify the impurities identified based on the parameters set in the model.

* **Output:** The Obtained results after data processing the model plots the graphical results showing various variations of parameters. The model also provides with water quality reports by predicting various parameter statuses.

**UML CLASS DIAGRAM WITH EXPLANATION**

**DATABASE DESIGN AS GRAPHICAL REPRESENTATION**





**Explanation of graphical representation of database design:**

We are designing a portal in a way that we can interact with the people around us through this.

* Users can raise their respected queries or put any complaints with respect to quality of water or any problems related to their water and mention their panchayat details.

So, the complaints are taken into consideration and actions are taken/provided by the officials accordingly. All the complaints raised are stored in a database table which is connected via our XAMP server. The data is stored in table containing the details of the users, location addressed, the problems they face.

* Admin logs in with their credentials and administrates the web portal data.
* User details contain personal details such as Name, Phone no, Date of birth and their location address.
* The location address consists of pin code, state name, district name, block name, village name, habitation name.
* The officials/ admins use this data to analyze the water quality in their areas and update the parameters into the obtained user recorded data from the complaints raised through the portal. The parameters are of various types, with different levels.
* The parameter types obtained are 5 which are Salinity, Fluoride, Iron, Arsenic, Nitrate.

The web portal database consists of user data, location details and parameters upon updating. Therefore 1 web portal requires a mandatory participation of the users to fill in their location details and personal details, officials to analyze the complaints and require a mandatory updating of the parameters. The final updates table is the dataset we use to assess, predict and manage the water quality.

The dataset is input into the model performing data processing and the obtained results after data processing, the model plots the graphical results showing various variations of parameters. The model also provides with water quality reports by predicting various parameter statuses.

**TOOLS USED**

1. **Weka (Dataset analysis)**:

Weka contains a collection of visualization tools and algorithms for data analysis and predictive modeling, together with graphical user interfaces for easy access to these functions. The original non-Java version of Weka was a Tcl/Tk front-end to (mostly third party) modeling algorithms implemented in other programming languages, plus data pre-processing utilities in C, and a Make file-based system for running machine learning experiments. This original version was primarily designed as a tool for analyzing data from agricultural domains, but the more recent fully Java-based version (Weka 3), for which development started in 1997, is now used in many different application areas, in particular for educational purposes and research.

**Advantages of Weka include:**

⮚ Free availability under the GNU General Public License.

⮚ Portability, since it is fully implemented in the Java programming language and thus runs on almost any modern computing platform.

⮚ A comprehensive collection of data preprocessing and modeling techniques. Ease of use due to its graphical user interfaces.

Weka supports several standard data mining tasks, more specifically, data preprocessing, clustering, classification, regression, visualization, and feature selection. All of Weka's techniques are predicated on the assumption that the data is available as one flat file or relation, where each data point is described by a fixed number of attributes (normally, numeric or nominal attributes, but some other attribute types are also supported). Weka provides access to SQL databases using Java Database Connectivity and can process the result returned by a database query. Weka provides access to deep learning with Deeplearning4j. It is not capable of multi-relational data mining, but there is separate software for converting a collection of linked database tables into a single table that is suitable for processing using Weka. Another important area that is currently not covered by the algorithms included in the Weka distribution is sequence modeling. The reason for using Weka in our project is mainly because it is easy to visualize the data from the data set and it is easy to use.

1. **Anaconda (Python and various python packages)**

Anaconda is a FREE enterprise-ready Python distribution for data analytics, processing, and scientific computing. Anaconda comes with Python 2.7 or Python 3.4 and 100+ cross-platform tested and optimized Python packages. All of the usual Python ecosystem tools work with Anaconda. Additionally, Anaconda can create custom environments that mix and match different Python versions (2.6, 2.7, 3.3 or 3.4) and other packages into isolated environments and easily switch between them using conda, our innovative multi-platform package manager for Python and other languages.

Reason: Basically, Anaconda has some list of advantages that I’ll list below,

● Depends on your preference though, but I prefer Anaconda because it gives the User ability to make an easy install of the version of python we want.

● You might not have an idea of how much a problem admin privilege can be if you are using your laptop and not a company owned. but Anaconda removes all those problems.

● It gives high performance computing with Anaconda Accelerate and several other components.

● Removes bottlenecks involved in installing the right packages while taking into considerations their compatibility with various other packages as might be encountered while using pip. There is no risk of messing up required system libraries.

1. **XAMP Server (Website and website testing):**

XAMPP is a free and open-source cross-platform web server solution stack package developed by Apache, consisting mainly of the Apache HTTP Server, MariaDB database, and interpreters for scripts written in the PHP and Perl programming languages. Since most actual web server deployments use the same components as XAMPP, it makes transitioning from a local test server to a live server possible.

**Advantages of XAMPP:**

✓ It is free and easy to use and easily available for Windows, Linux and Mac OS.

✓ It is a beginners friendly solution package for full stack web development.

✓ It is a open source software package which gives a easy installation experience.

✓ It is very simple and lightweight to create set up for development, testing and deployment.

✓ It is a time-saver and provides several ways for managing configuration changes.

**Test Cases**

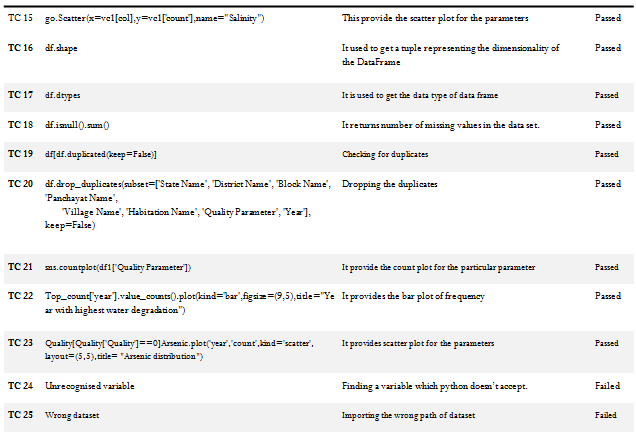
**Test Cases obtained from python scripting using Anaconda:**

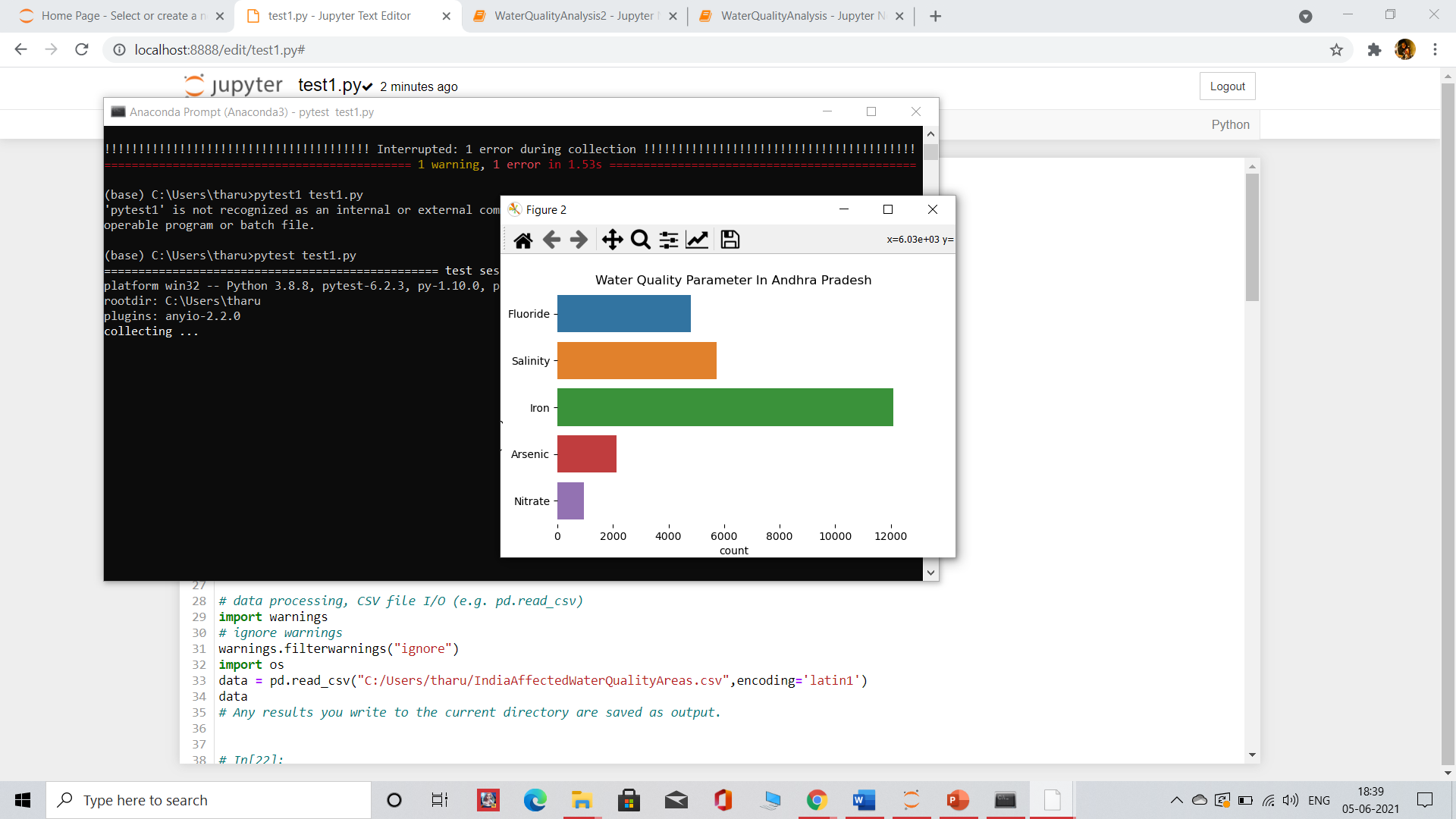
We used pytest package to test our scripts

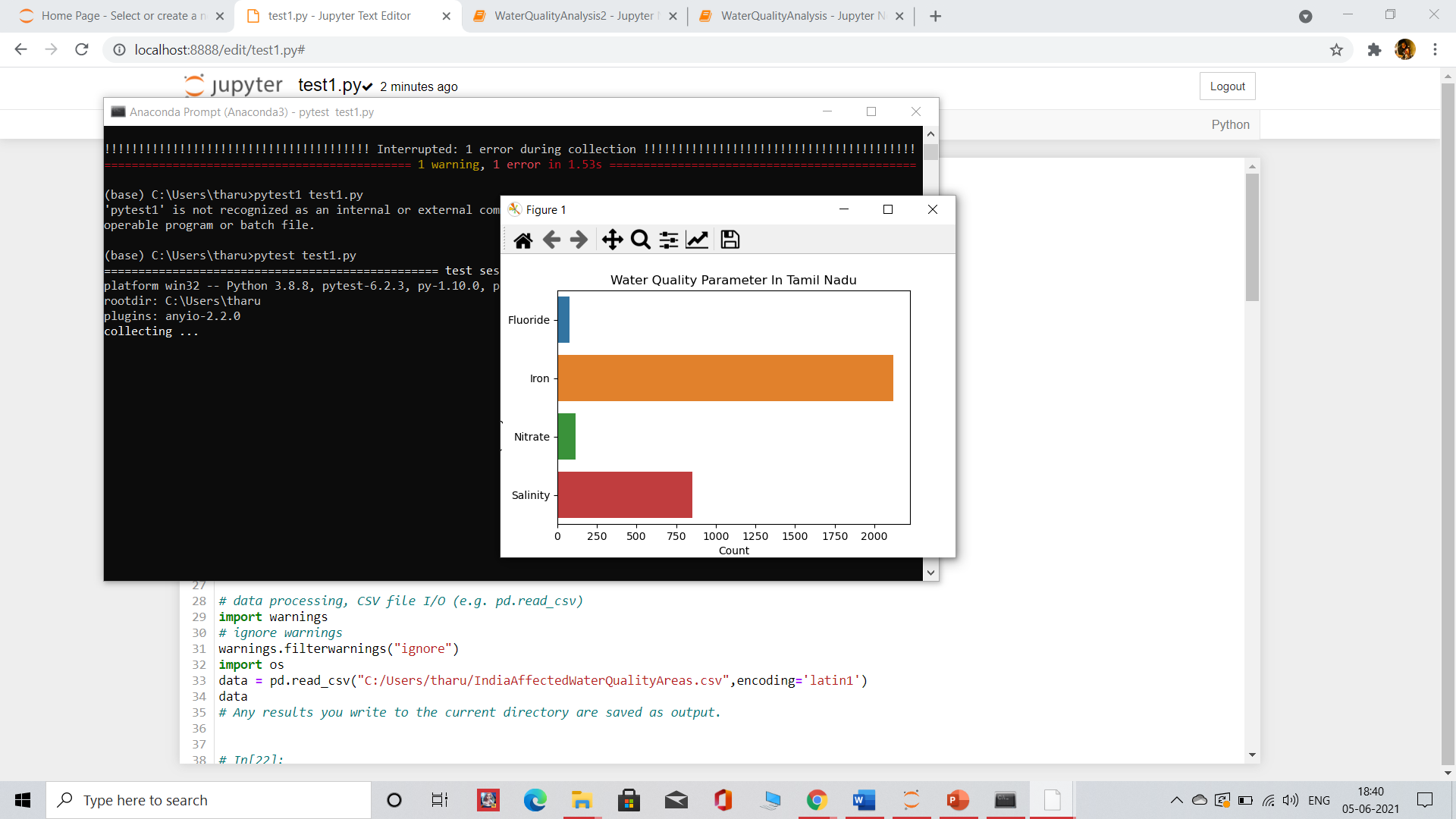
## 

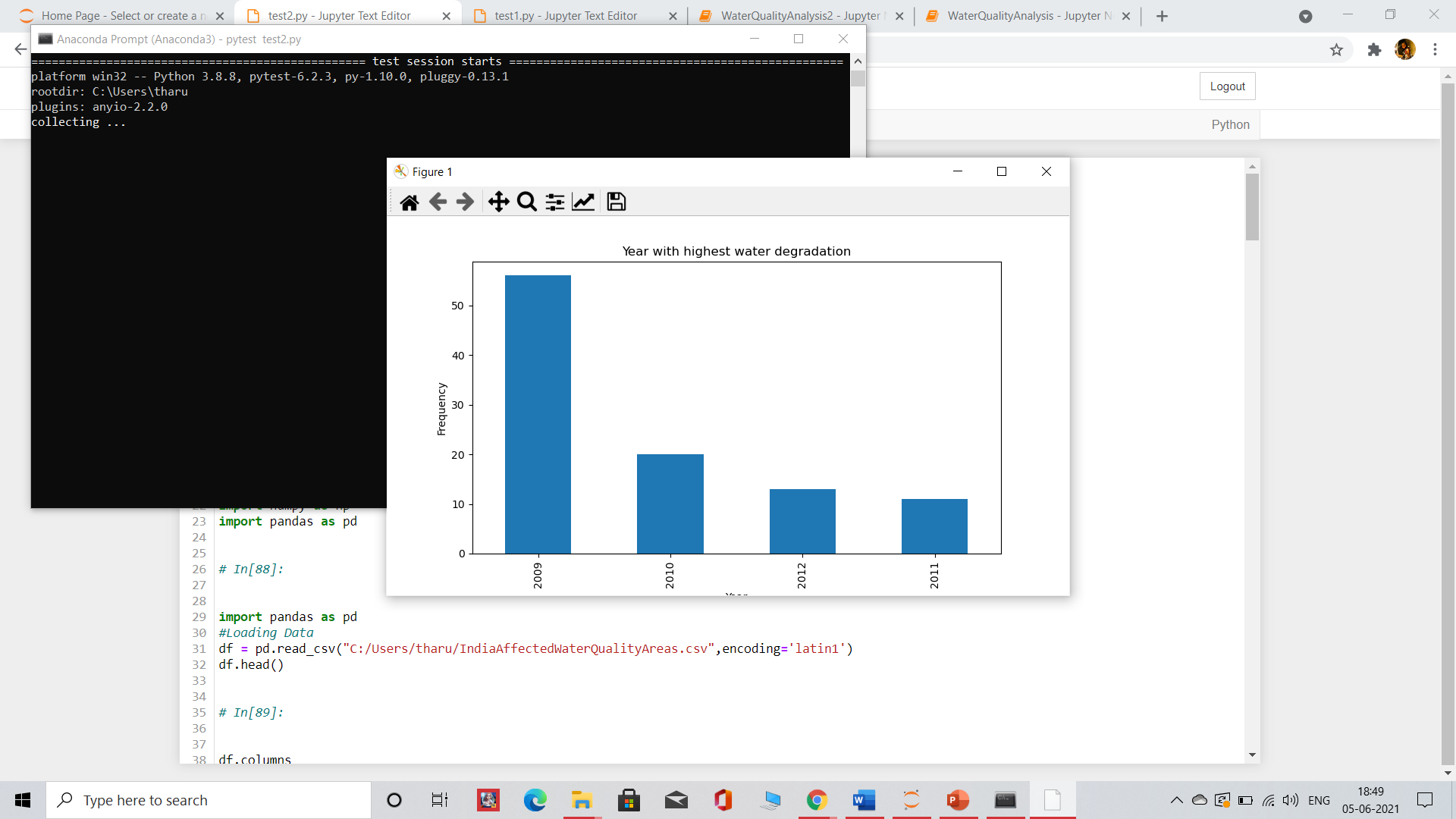
## 



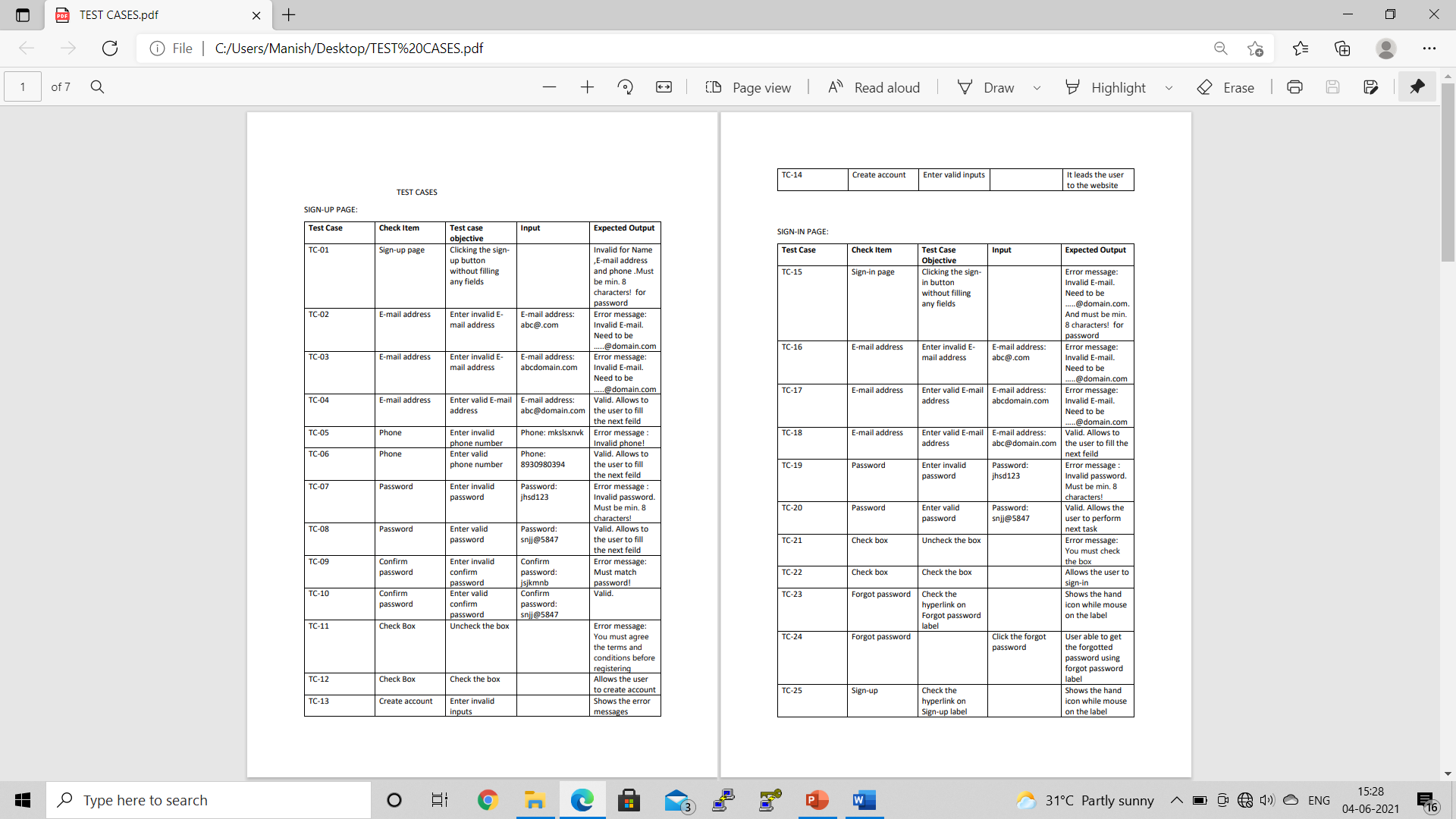


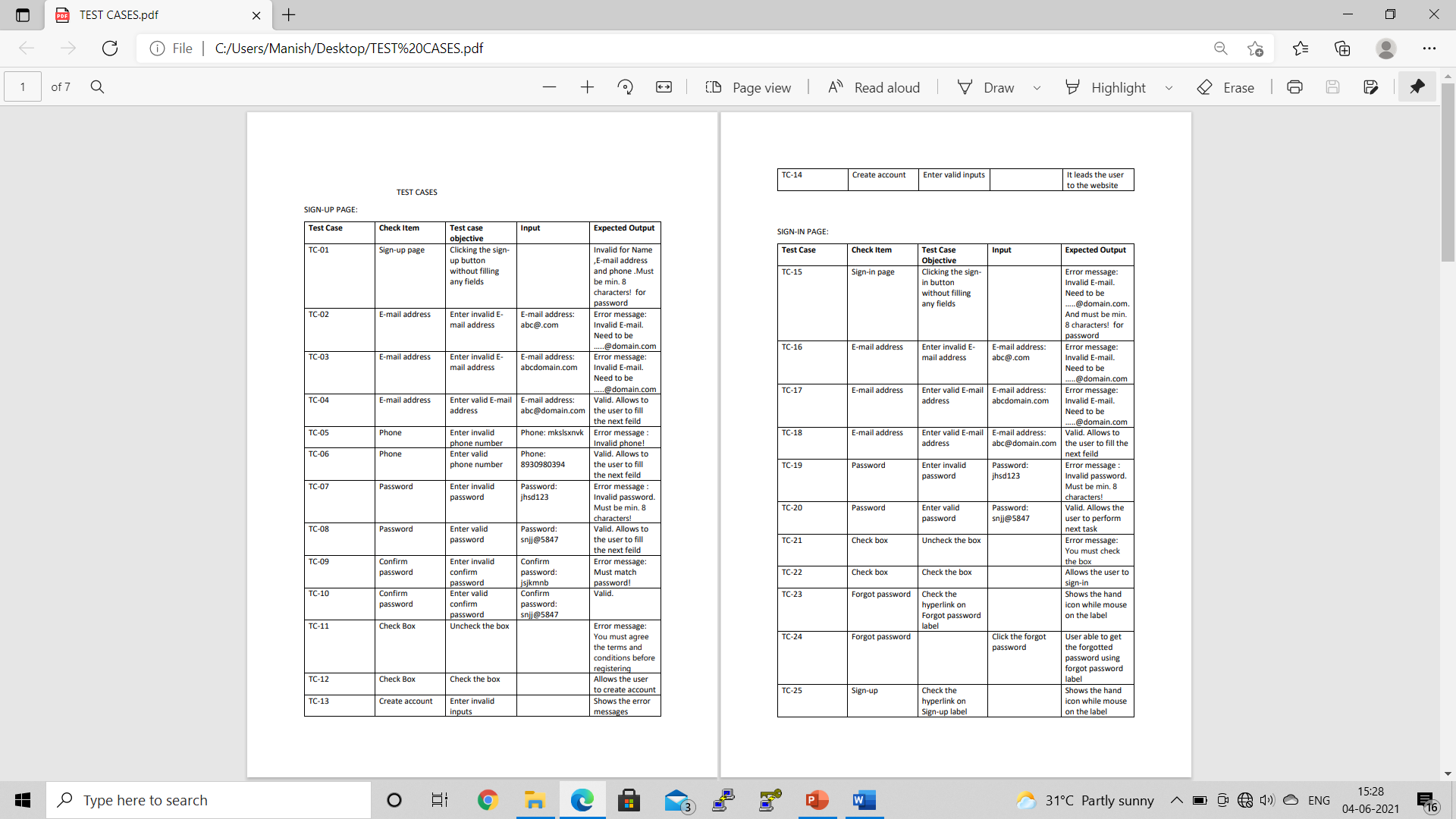


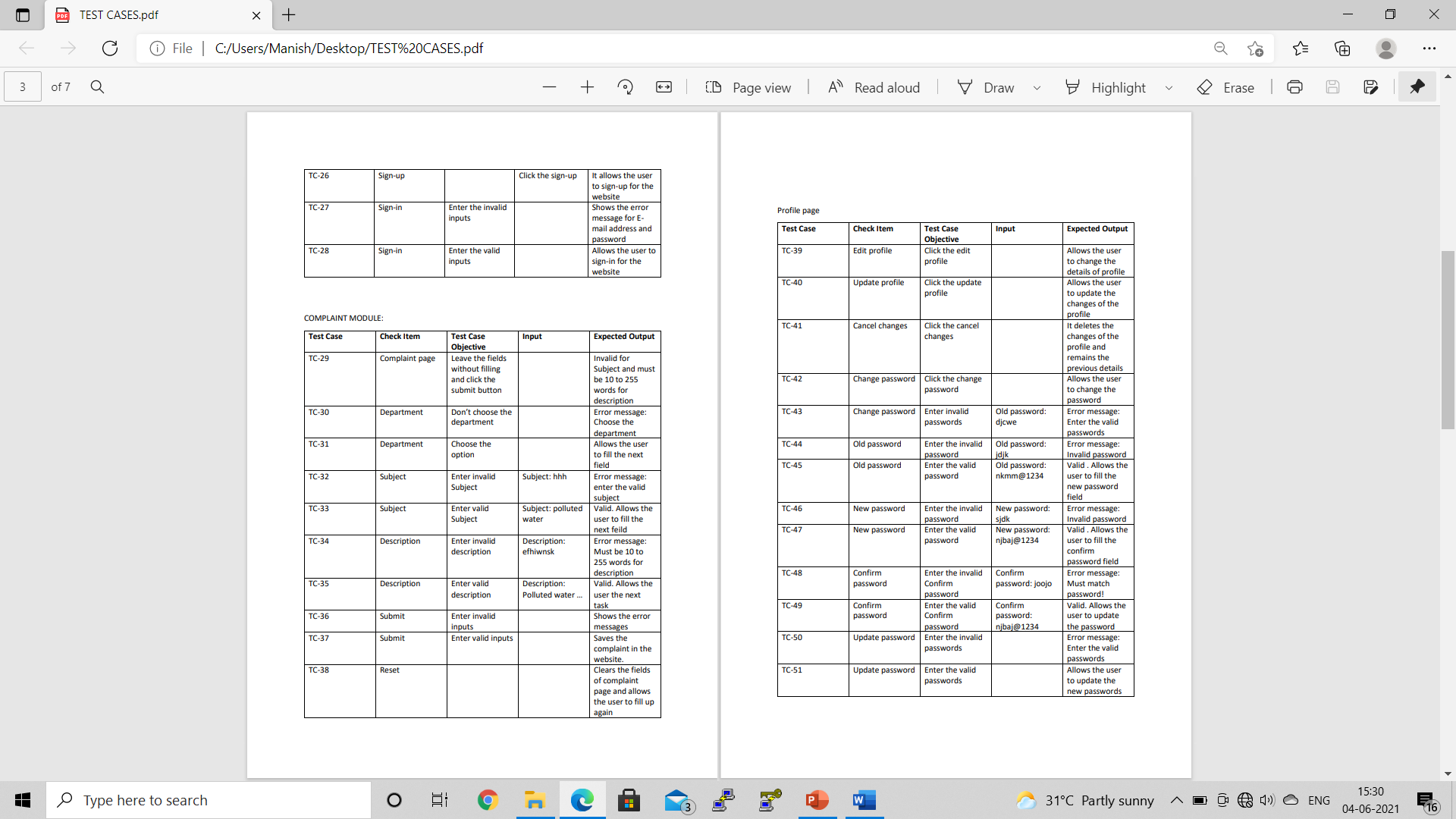


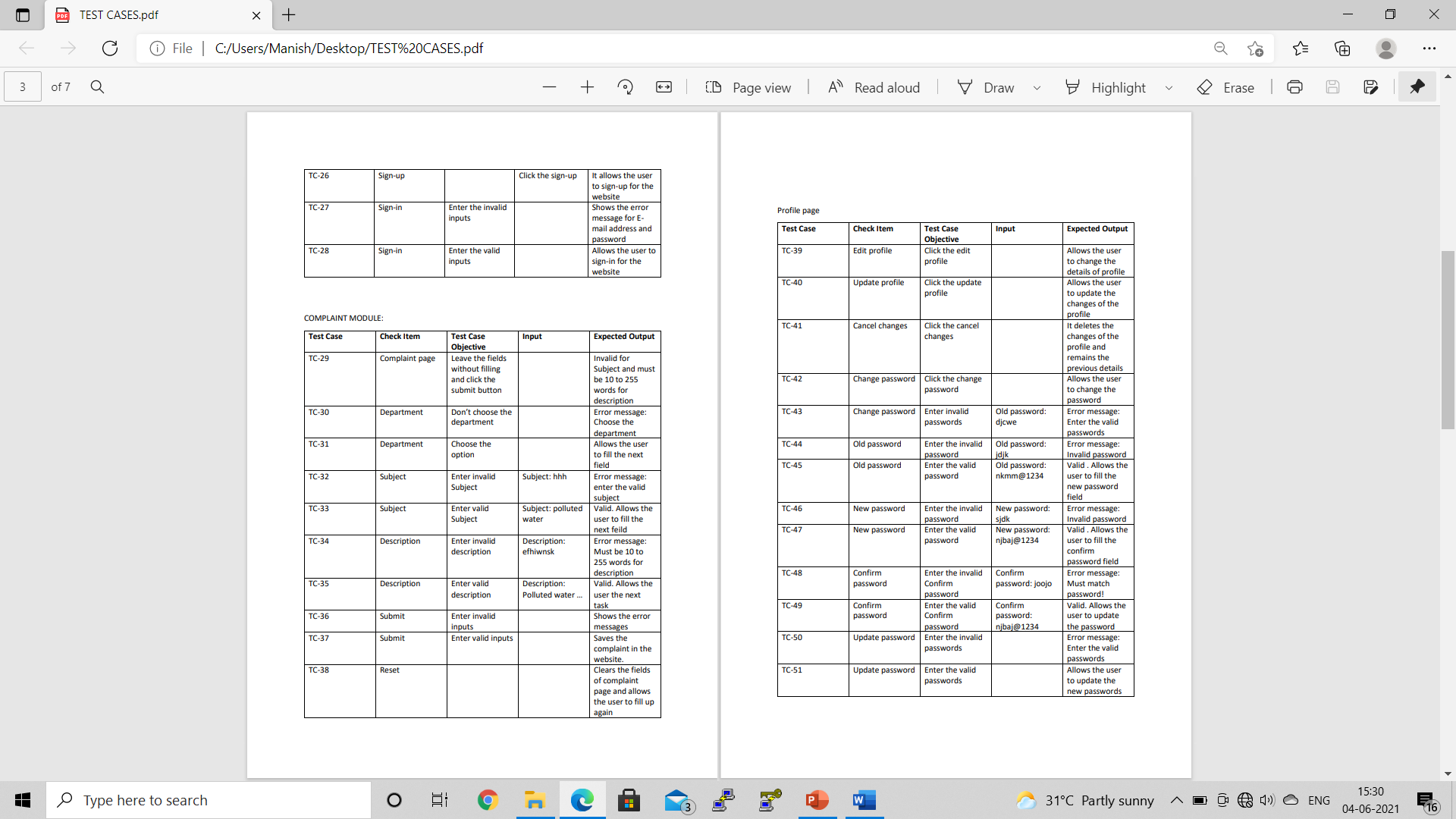


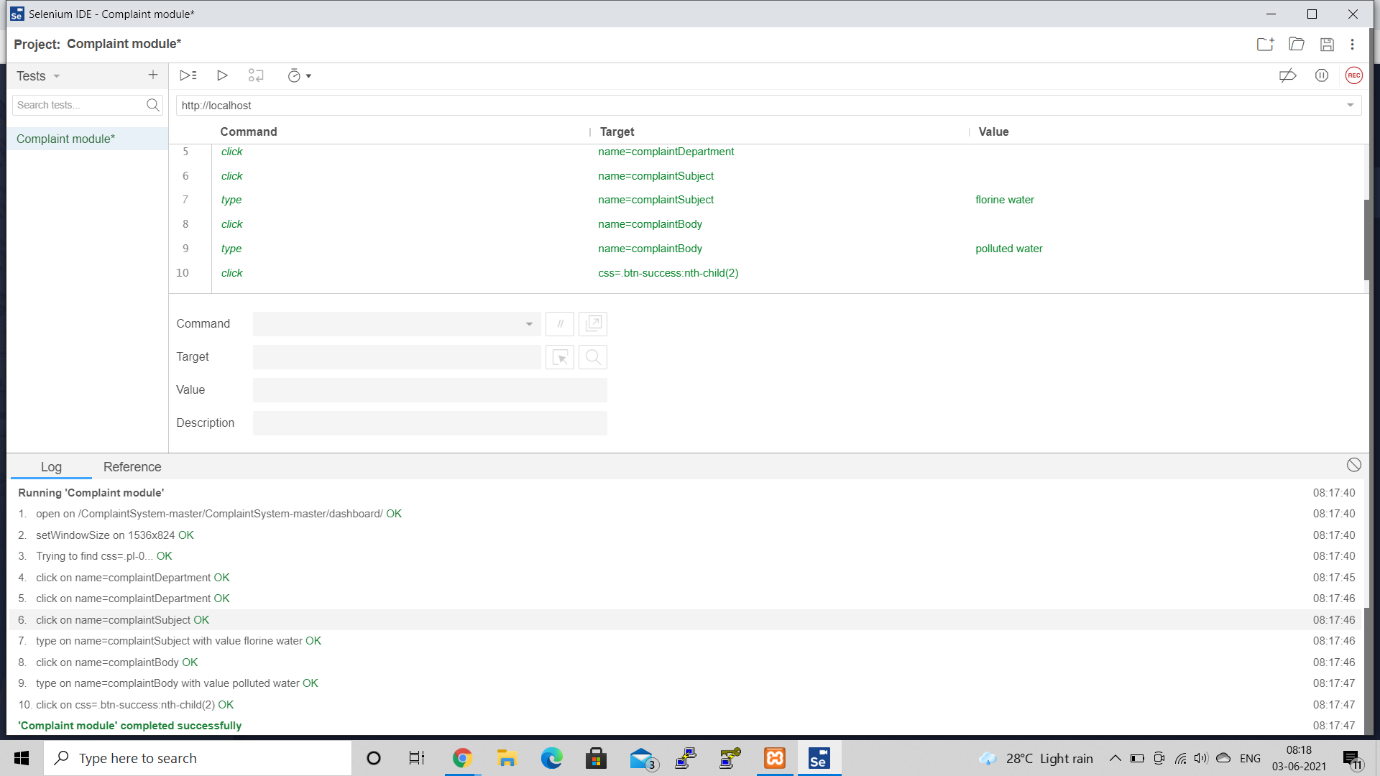
**Testcases obtained by using selenium tool on our website frontend and backend:**

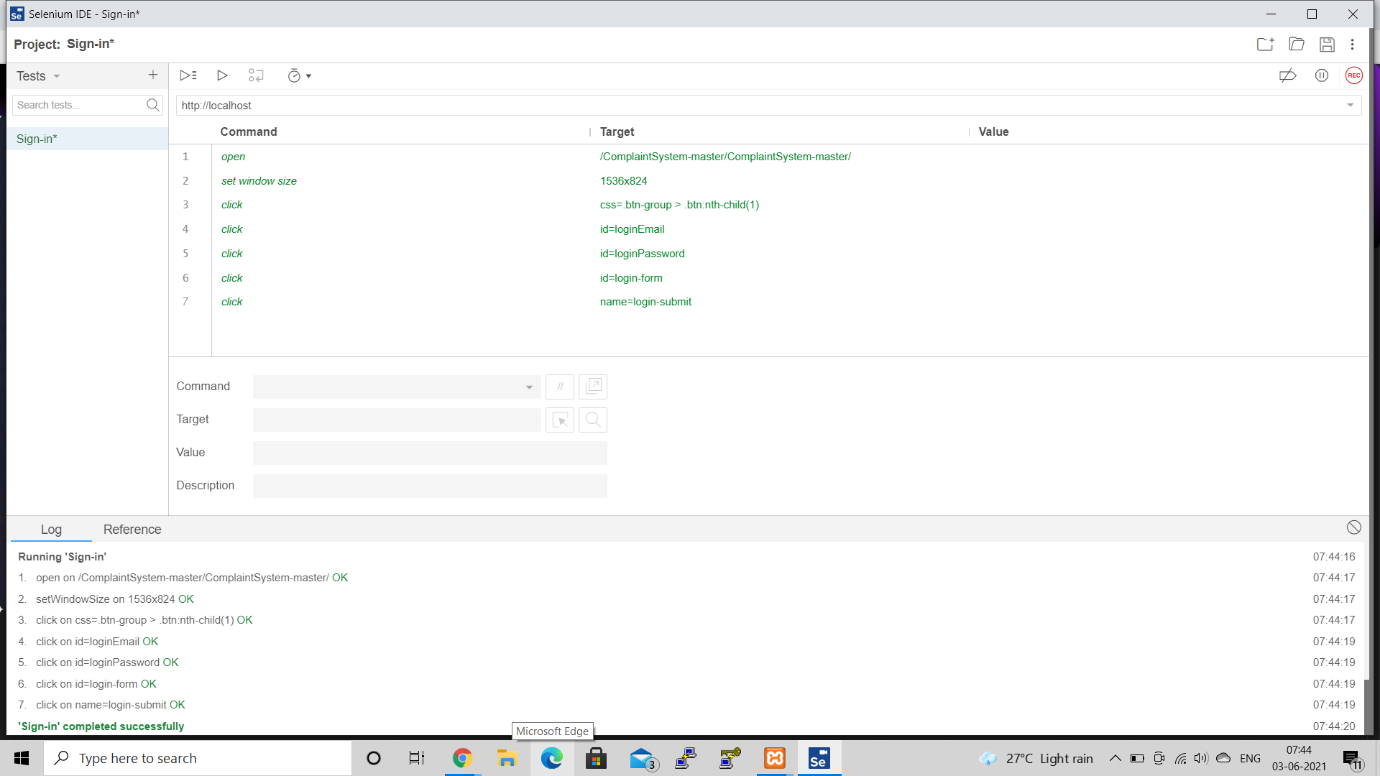


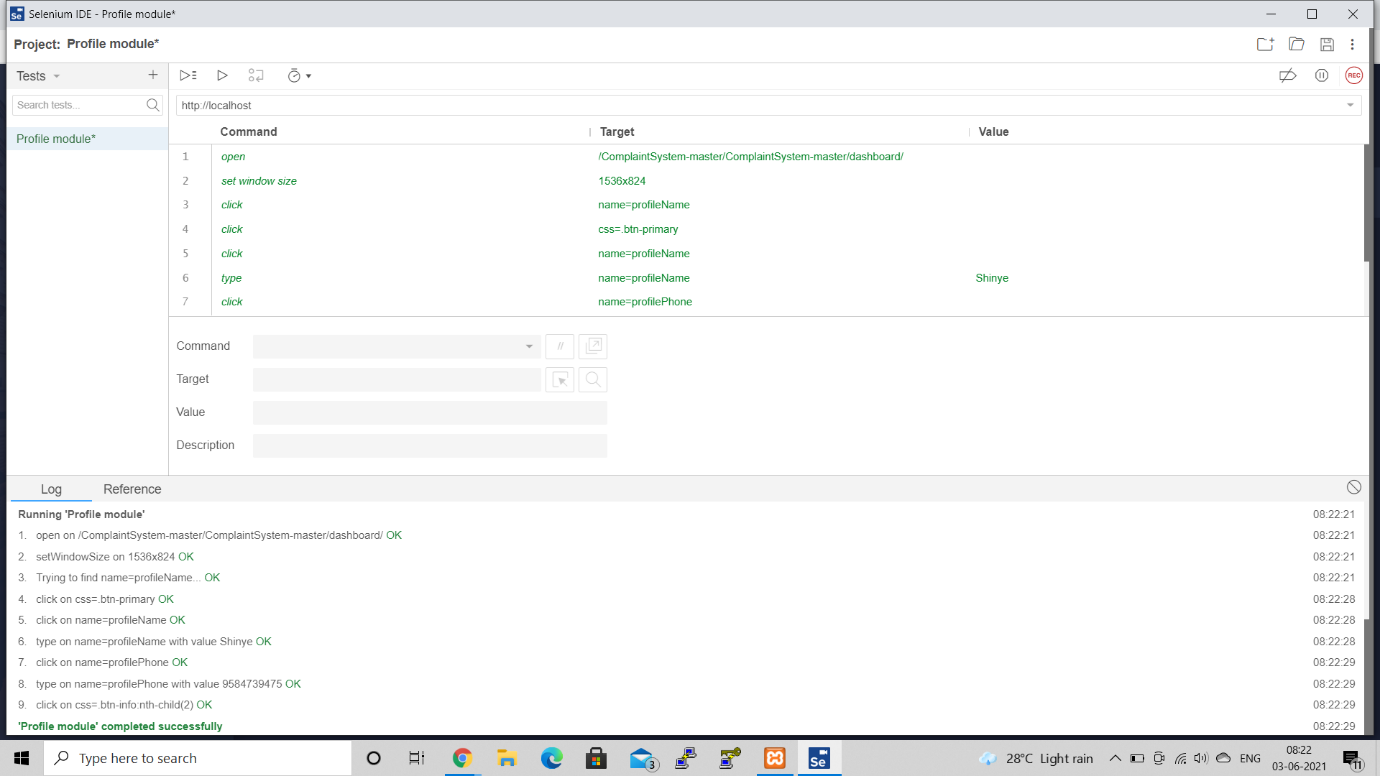




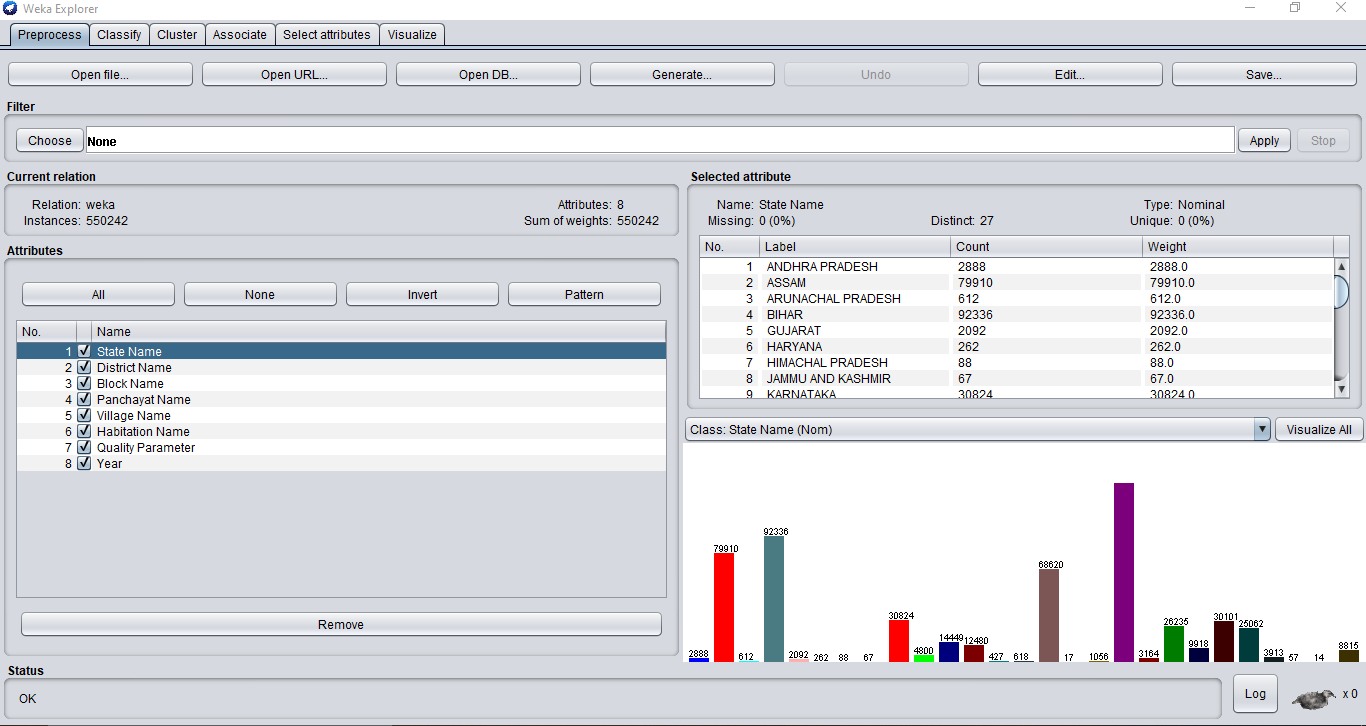


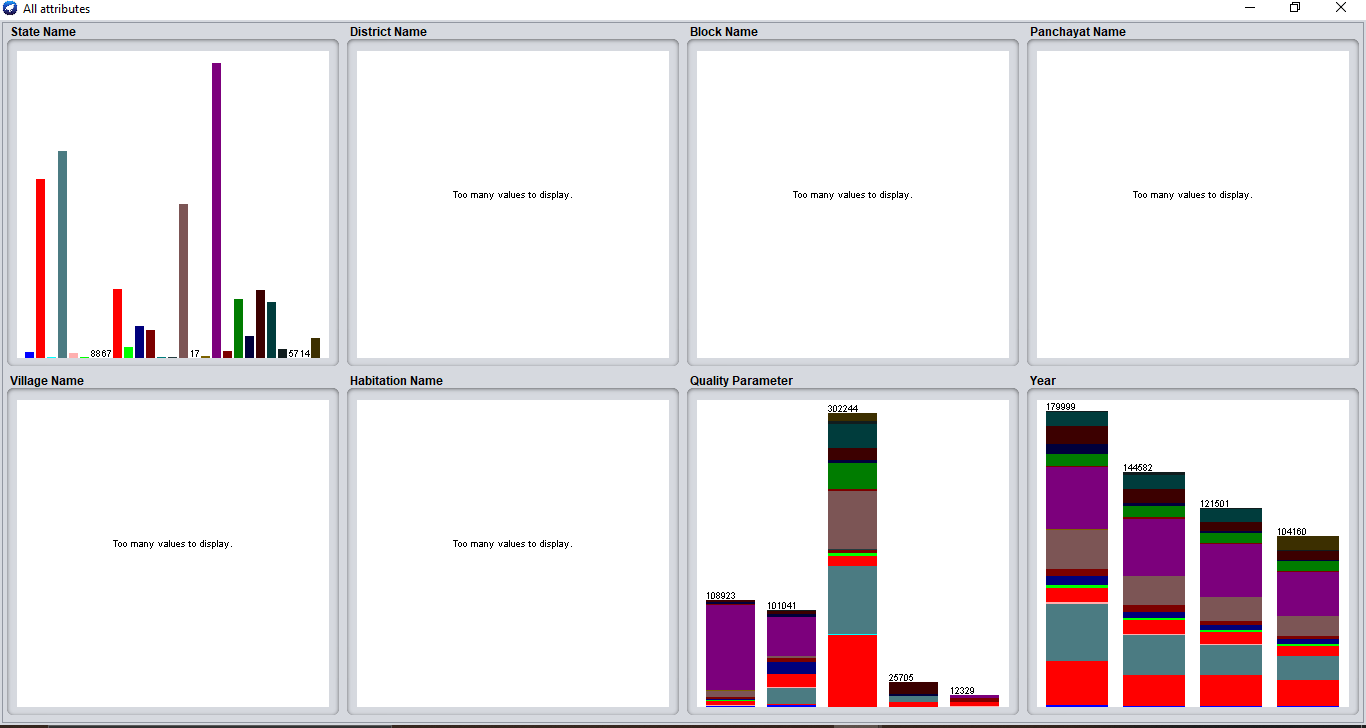


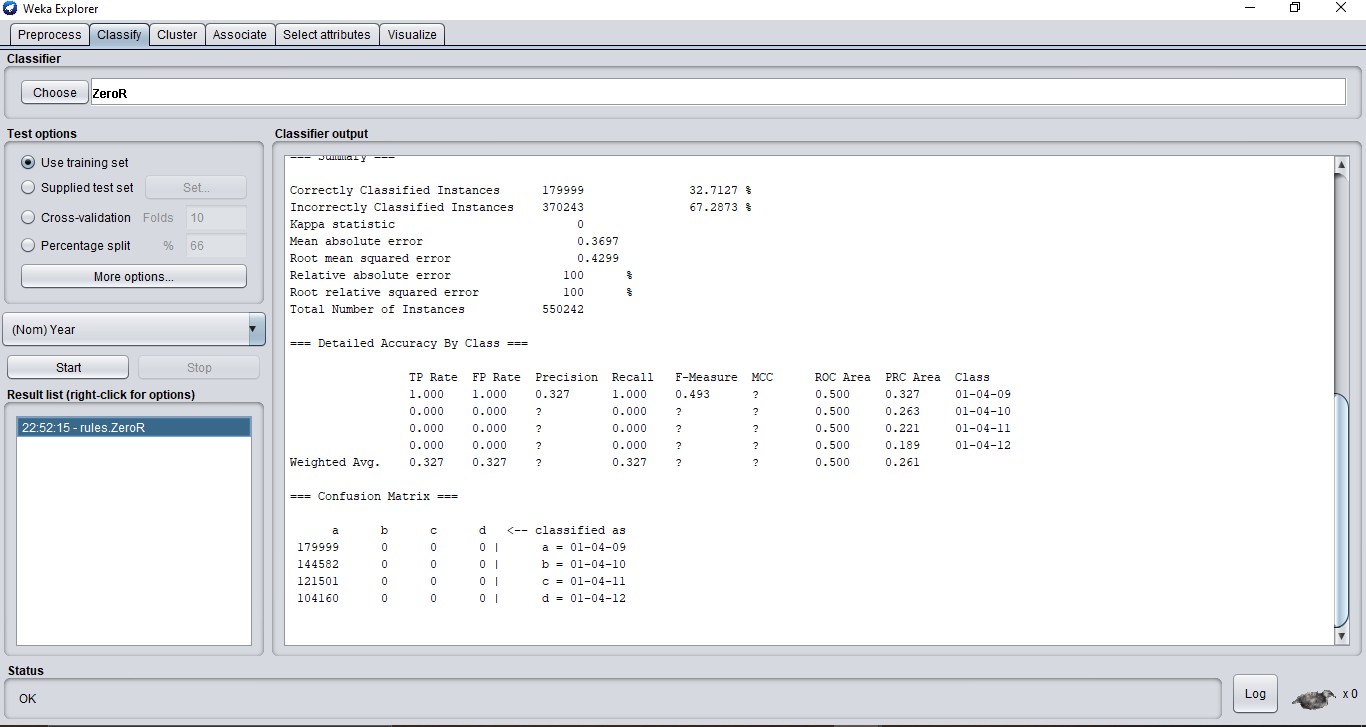


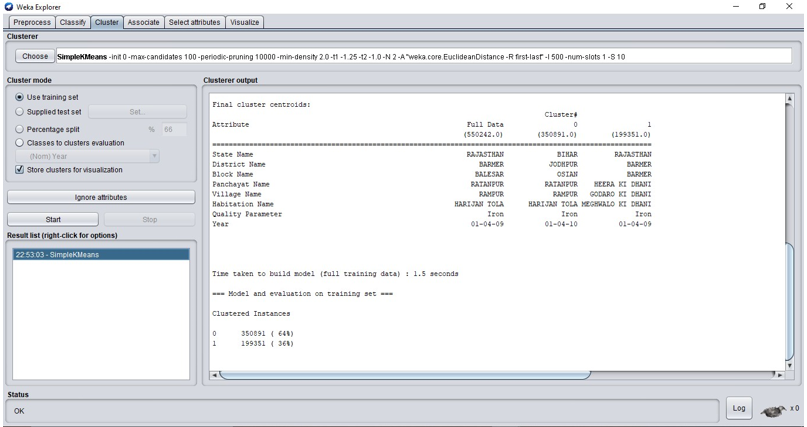


**Visualization using WEKA tool:**









**CONCLUSION**

The project helps people register complaints, know the status of water in their area. The obtained results after data processing, the model plots the graphical results showing various variations of parameters. The model also provides with water quality reports by predicting various parameter statuses.

With the help of python programming, we have achieved predictive analysis of water quality in various parts of India. This project fulfills the partial requirement of government portal.

Water Quality management has become a major problem of overall water resources management. The people who are living in rural areas will depend on low levels of water quality. That poor people don’t know what the chemicals contained in that low-level water. More pollutants will be generated if the populations and economics grow and finally, they end up with the surface and groundwater bodies. Every living thing or organism requires water not for just sufficient quantities but also contains quality to survive healthy. The main aim of this project is to identify the values, which are more affected by scanning the dataset and taking care to reduce contaminants in that rural area. Drinking water quality can be determined by performing water quality analysis, which is a major component of the water distribution system.

Time and Studies comparing classification algorithms have found that the Naïve Bayesian classifier has always been comparable in performance with decision tree and selected neural network classifiers. Besides, they have also exhibited high accuracy and speed when applied to large databases. Even our simulation results have proven this special classifier to be optimal and efficient in successfully working on mined datasets.

The analysis system can be improved by constant updating of the quality parameters on a regular basis, implementing various other machine learning algorithms for greater efficiency, introducing cloud-based storage services to prevent the server from crashing and better traffic control, implementing python code for segregation of water related problems Soil pollution parameters, Industrial pollution parameters.

**REFERENCES**

[1] Rothwell, J.J., Dise, N.B., Taylor, K.G., Allott, T.E.H., Scholefield, P., Davies, H., Neal, C., Predicting river water quality across North West England using catchment characteristics, Journal of Hydrology (2010), doi: 10.1016/j.jhydrol.2010.10.015

[2] C. R.Ramakrishnaiah, C. Sadashivaiah, and G. Ranganna, “Assessment of Water Quality Index for the Groundwater in Tumkur Taluk, Karnataka State, India,” E- Journal of Chemistry, vol. 6, no. 2, pp. 523-530, 2009.

[3] Yidana, S.M. &Yidana, A. Environ Earth Sci (2010) 59: 1461.

[4] Feng Han and Yi Zheng, Joint analysis of input and parametric uncertainties in watershed water quality modeling: A formal Bayesian approach, Advances in Water Resources, 10.1016/j.advwatres.2018.04.006, 116, (77-94), (2018).

[5] William Ocampo-Duque, NúriaFerré-Huguet, José L. Domingo, Marta Schuhmacher, Assessing water quality in rivers with fuzzy inference systems: A case study,EnvironmentInternational,Volume 32, Issue 6,2006,Pages 733 742,ISSN 0160- 4120.

[6] Gremillion, P. T. & R. F. Malone, 1986. Waterfowl waste as a source of nutrient enrichment in two urban hypereutrophic lakes. In Lake and Reservoir Management, Vol. 2. Proceedings of the Fifth Annual Conference and International Symposium on Applied Lake.

[7] Barkman, J. H., & Davidson, D. H. (1972, July 1). Measuring Water Quality and Predicting Well Impairment. Society of Petroleum Engineers. doi:10.2118/3543-PA

[8] Holger R Maier, Nicolas Morgan, Christopher W.K Chow,Use of artificial neural networks for predicting optimal alum doses and treated water quality parameters,Environmental Modelling &Software,Volume 19, Issue 5,2004,Pages 485 - 494,ISSN 1364-8152.

[9]Julia A. Howitt, Darren S. Baldwin, Gavin N. Rees, Janice L. Williams,Modelling blackwater: Predicting water quality during flooding of lowland river forests,EcologicalModelling,Volume 203, Issues 3–4,2007,Pages 229-242,ISSN 0304- 3800.