Arab American University

Faculty of Graduate Studies

Predicting the effects of weather conditions on agriculture by applying spatial data mining methods

By

Mohammed Omer Qasem Eleyat

Supervisor

Dr. Jacqueleen Joubran Abu Daoud

This thesis was submitted in partial fulfillment of the requirements for the Master`s degree in

Computer Science

2018

Predicting the effects of weather conditions on agriculture by applying spatial data mining methods

By

Mohammed Omer Qasem Eleyat

This thesis was defended successfully on 24/1/2018 and approved by:

Committee members	signature
1.Supervisor name: Dr. Jacqueleen Joubran Abu Daoud	…….
2.Internal Examiner Name: Dr. Amjad Rattrout	…….
3.External Examiner Name: Dr. Labib Arafeh	…….

Declaration

This is to declare that the thesis entitled " Predicting the effects of weather conditions on agriculture by applying spatial data mining methods " under the supervision of Dr. Jacqueleen Joubran Abu Daoud is my own work and does not contain any unacknowledged work or material previously published or written by another person, except where due reference is made in the text of the document.

Dedication

This thesis is dedicated to my best friends who have always been a constant source of support and encouragement during the challenges of my whole college life. Also to my brothers and Sisters whom am truly grateful for having you my life. This work is also dedicated to my parents, who were in love to me unconditionally and whose good examples have taught me to work hard for things that I aspire to achieve.

Acknowledgements

I would like to express my gratitude to my supervisor Dr. Jacqueleen Joubran Abu Daoud, Chair of the GIS Department in the Faculty of Engineering and Information Technology, Arab American University, Palestine for her keen interest guidance, patience and assistance throughout this research work.

My special thanks to officials of Ministry of Agriculture and Meteorological Station for giving me the data about weather conditions and percentage of affected crops from 2010 to 2015.

I am highly thankful to all of my teachers and friends and surrounding people who assist and encourage me to complete this research.

Abstract

Weather conditions including precipitation, temperature, wind speed, solar radiation, and humidity affect plants in different ways that can make them more susceptible to disease and insect problems. Unlike the short term weather, the climate represents the average weather conditions over a long period of time, which determines what will probably grow well in a certain region. However, extreme weather conditions can kill plants and damage whole farms, which may result in huge agricultural loss.

Due to the importance of the weather forecast to the scientific and technological issues and challenges of the century problems, the researcher applied spatial data mining methods to the collected weather data in order to forecast weather conditions and alert farmers to take precautions and avoid agricultural damage. Specifically, with reference to the Palestinian Meteorological Authority and Ministry of Agriculture, the collected data included statistics about precipitation, temperature, wind speed, solar radiation, humidity and the percentage of the affected crops.

These data were digitized, prepared, cleaned and normalized to make ensure that the analysis and results are correct. Then, they were converted into GIS formats and ArcGIS. After that, the spatial data of the research were saved, visualized, joined and analyzed using GIS software. The data were also interpolated i.e. data were extended so that they cover all the points that are around 1000 meters apart from each other. Several algorithms, such as ordinary least square and multi perceptron neural network of data mining, were implemented to predict the percentage of the affected crops before and after interpolation. Some of these algorithms were applied to ArcGIS software and some of them to the WEKA software; some codes were programmed to convert data, sort them and apply the new spatial data mining implementations.

Consequently, a new approach is suggested for spatial prediction by using map algebra and dealing with the maps area as a matrix. This spatial approach was applied taking into consideration the surrounding four neighbors for each location point to ensure the inclusion of the effect of these surrounding areas, fields, properties and effects.

All the methods were applied to the training data and tested by cross validation. The adjusted residuals were also calculated and compared. In conclusion, The results of each implementation were examined. The best results were achieved using the neural network method with 0.0718 mean absolute error ,0.1664 root mean squared error and 0.3714% relative absolute error. Residuals of testing data and cross validation were minimized and compared. Satisfying results were achieved and presented. A clear improvement was achieved to here, The mean absolute error is reduced from 1.3837 to 0.0718, by the suggested spatial model in three levels when spatial neighbor's variables were considered. Matrix map algebra was done and furthermore the target values themselves of the neighbor areas were considered using iterations. The results of the iterations have been tested by calculating the maximum difference between the current and previous iteration. Consequently an impressive improvement was achieved as the K maximum error is reduced from 14.377 to 1.251, for the number of iterations.

This, in turn, encourages more research in this field and suggests future work that includes advanced analysis and modeling.

Table of Contents

1. Introduction....1
2. Problem statement....5
3. Motivation....6
4. Contribution....7
5. Background and Literature Review....8
5.1 Geographic Information System (GIS)....8
5.2 Data mining....8
5.3 Neural network....11
5.3.1 Structure of Neurons in Brain....12
5.3.2 Advanced Neural Network (ANN) architectures....13
5.3.3 The most Common Types of Learning in Neural Network....16
5.3.4 Learning Data Sets in ANN....16
5.3.5 Four Different Uses of Neural Networks....17
5.4 Normalization....18
5.5 Interpolation....19
5.6 Space Time Cube....20
5.7 Cross - validation and multilayer perceptron network....21
5.8 Weka....23
5.9 Related Work....23
6. Experimental Setup and Work....29
6.1 Data collection and preparation.....29
6.2 Georeferencing Map and Join data to the ArcGIS map.....31
6.3 Data interpolations....32
6.4Applying several algorithm for percentage of affected crops prediction....34
6.4.1 Ordinary least square....34
6.4.2 Applying multilayer perceptron neural network using Weka....35
6.5 Suggested Improvement Method of spatial analysis and data mining....37
7. Results and Discussions....42
8. Conclusion and Future Work....65
9. References....69
10. ملخص الرسالة باللغة العربية....75

List of tables

Table 1: the summery of the related work..... 27

Table 2: comparison between the results of the Ordinary Least Square with or without elevation and slope.....59

Tabel 3: The results of prediction for all different technique 64

List of figures

Figure 1- Data Mining Techniques[15]. 11

Figure 2- The typical nerve cell of human brain comprises of four parts[16] 12

Figure 3- Single Layer Feed Forward Network[17] 13

Figure 4- Architecture of Neural Network with Hidden Layers[17] 14

Figure 5- RBF Network Architecture[19] 15

Figure 6- Neural Network for Pattern Recognition 17

Figure 7- Space time cube (A time series as space-time bins). [8] 21

Figure 8- Jenin district our study area[39] 30

Figure 9- Collecting information and converting raw data to excel sheet 31

Figure 10 - join xls data file with areas in jenin area 32

Figure 11- 850 new point for each attribute every month after interpolation. 33

Figure 12- applying ordinary least square on our data 35

Figure 13 - our model in prediction in multi perceptron neural network using 2 hidden layer. 36

Figure 14 - Results with summary of statistics console. 37

Figure 15- - Using the neighbors for each point over time 38

Figure 16 - The option of neighbors and their location as can be considered in the spatial 39

Figure 17 - using K iteration of the target values (affected percentage) to get more accurate results with the neighbors of the target value in iteration k+1. 41

Figure 18 - shows IDW classes that represents weather variables effects in our study area 42

Figure 19 - IDW Rain raster in April / 2015 42

Figure 20- IDW Rain raster in March / 2015 43

Figure 21- IDW Rain raster in February / 2015 43

Figure 22- Number of Affected items in February / 2015 43

Figure 23- Affected Percentage in February / 2015 44

Figure 24- Number of Affected persons in February / 2015 44
Figure 25-IDW rain raster in January / 2015 44
Figure 26- IDW rain raster in December / 2014 45
Figure 27- IDW rain raster in November / 2014 45
Figure 28- IDW rain raster in October / 2014 46
Figure 29 - IDW rain raster in May / 2014..... 46
Figure 30 - IDW rain raster in march / 2014..... 46
Figure 31 - IDW number of affected items raster in march / 2014 47
Figure 32 - IDW Affected percentage raster in March / 2014 47
Figure 33 - IDW number of affected persons raster March / 2014 47
Figure 34 - IDW rain raster in February / 2014 48
Figure 35 - IDW Rain raster in January / 2014..... 48
Figure 36 - IDW rain raster in December / 2013..... 49
Figure 37 - IDW rain raster in November / 2013 49
Figure 38 - IDW rain raster in April / 2013..... 49
Figure 39 - IDW rain raster in March / 2013 50
Figure 40 - IDW rain raster in February / 2013 50
Figure 41 - IDW rain raster in January / 2013..... 50
Figure 42 - IDW number of affected items in January / 2013..... 51
Figure 43 - IDW Affected percentage raster in January / 2013 51
Figure 44 - IDW number of affected persons in January / 2013..... 51
Figure 45- IDW Rain raster in December / 2012 52
Figure 46- IDW rain raster in November / 2012 52
Figure 47- IDW rain raster in October / 2012 53
Figure 48- IDW rain raster in March / 2012 53
Figure 49- IDW rain raster in February / 2012 53
Figure 50- IDW number of affected items in February / 2012..... 54

Figure 51- IDW Affected percentage raster in February / 2012 54

Figure 52- IDW number of affected persons raster in February / 2012..... 54

Figure 53- IDW rain raster in January / 2012 55

Figure 54- Converting topographic contours into 3D surface and raster of Jenin elevations (the image with z factor for elevations) 56

Figure 55- Raster output file that shows the calculation of the slope of the topographic surface in each point..... 56

Figure 56- A summary of OLS results – Model variables without elevation and slope 57

Figure 57- summary of OLS results – Model variables..... 57

Figure 58 - summary of OLS results – Model variables 58

Figure 59- residuals with elevations and with slope..... 59

Figure 60 - variable distributions and relationships 60

Figure 61 - graph of error percentage between standard residuals and predicted values in 2013 and 2014 61

Figure 62 - Applying K-fold Cross validation and multilayer perceptron network for data before interpolation..... 62

Figure 63 - Applying K-fold Cross validation and multilayer perceptron network for data after interpolation. 63

Figure 64 -The results of aspect analysis on Jenin 3D surface..... 67

List of abbreviations

GIS -Geographic Information System

ANN -Advanced Neural Network

BPN - Back-propagation Network

RBF -Radial Basis Function Network

ANFIS -Adaptive neuro-fuzzy inference system

IDW-Inverse distance weighted

TSA - Trend surface analysis

LOO-XVE - leave-one-out cross validation error

IDAMS - Integrated Data Analysis and Simulation Module

AMS - Analysis and Simulation Module

BOVIS - Bovine Information System

OLS - Ordinary least square

1. Introduction

Agriculture plays an important role in the economy of any country as it provides food, raw materials, and job opportunities. However, people working in agriculture in the developing countries are typically much poorer than those working in the other sectors due to the loss of crops as a result of the natural disasters (snow and frost, hurricanes, floods, fires and tornadoes).

Geographic Information System (GIS) is a system that collects, inputs, processes, analyzes and outputs spatial data. This system also assists planning and decision-making in several fields related to agriculture, urban planning, housing expansion, as well as reading the infrastructure of any city through the creation of so-called Layers. This system integrates the geographical information input (maps, aerial photographs, and satellite images), descriptors (names, tables) and their processing (revision of the error) stores, retrieves, undergoes the spatial and statistical analysis, and displays the results on the computer screen.

The availability of spatial data in the form of vast and high resolution has provided opportunities to acquire new knowledge, extract features, get a better understanding of the complex geographic problems, such as human-environmental interaction, socio-economic dynamics and address pressing several problems, such as the climate change and the spread of pandemic influenza.

Data mining is the process of identifying and creating pattern relationships by sorting out large data sets to solve problems through data analysis. Data mining is also used to predict future by using tools for modeling and analyzing the autocorrelation between the considered variables.

Spatial data mining is a kind of data mining which involves the process of discovering interesting unknown patterns from spatial dataset. Spatial data, also known as geospatial data, is information about location that can be represented as value numbers in a geographic coordinate system. Generally, spatial data represents the location, size and the shape of an object on the Earth surface, such as a building, mountain and towers. Spatial data may also include variables that have information about the object, such as elevation, slope and pressure that is represented.

Extracting knowledge and useful patterns from database is very difficult because of the complexity of data types, relationship and autocorrelation. Preprocessing is also a very important step for spatial data mining to deal with problem , such as missing location information, cleaning, feature selection, and data transformation. [1]

Recently, Geographic Information System (GIS) has been used not only to capture, store and retrieve geospatial data, but also to query, analyze data and visualize results in the form of maps. Existing tools, such as map creation, overlay, classification, etc. are available in GIS tools with the ability to be used as a spatial data analysis tool. However, even modern GIS provides limited tools for analyzing complex spatial data and discovering knowledge and does not have sufficient capacity to integrate decision maker preferences, experiences, intuition, and judgments into the problem resolution process [2].

Similarly, spatial data extraction and geographic information systems alone provide limited methods for dealing with uncertainty in spatial data. Therefore, geographic information systems have limitations on spatial data features, which must be determined in advance. Thus soft computing techniques including neural network can be used to solve the above issues. The huge explosion of geographically referred to site-sensitive data, technological advances, remote sensing and digital mapping

combined with the need to address uncertainty in spatial data and the use of linguistic terms, underlines the importance of integrating spatial data mining، GIS, neural network [3][4].

Meteorological data mining can be used to find hidden patterns within the available meteorological data and retrieve the information leading to useful knowledge. This can play an important role in analyzing and extracting the climate variability and climate prediction, which can be useful in supporting many important fields, such as agriculture, plants, and water resources [5].

The anticipated outcome of this research is helping farmers reduce their risks by protecting their farms from diseases and other agricultural dangers in an attempt to increase their incomes by predicting the percentage of affected crops based on the data collected from the meteorological sites and the Ministry of Agriculture.

In this research, the researcher collected necessary weather data including precipitation, temperature, wind speed, solar radiation, and humidity from the meteorological sites and the Ministry of Agriculture and then applied data mining methods to forecast weather conditions and alert farmers based on the prediction of agricultural risks to minimize losses. Raw data was collected from the Palestinian Meteorological department, the Palestinian Ministry of Agriculture, and from online sources.

Because weather data can be used to predict plants diseases and insect problems , so that farmers can take precautions and safety measures to protect their farms from diseases or death, the researcher will use these data to predict the anticipated dangers against agriculture to help the farmers protect their plants.

This thesis is organized as follows: Chapter one includes an introduction about the main concepts of our work; chapter two and three define and explain the problem and motivation; chapter four introduces the main contributions of this work , while background and related work are explained in chapter 5. Chapter 6 is about the methodology used to meet the thesis goals. After that, the results are presented and discussed in chapter 7. Finally, the thesis is ended with a conclusion and future work.

2. Problem statement

Agriculture is the most important economic field and the rural poor in the developing countries will be more vulnerable to the effects of weather conditions. The current policies regarding weather conditions mitigation and adaptation require interventions at different levels of research ranging from crop management and guidance to the farmers to protect their crops.

In view of the unpredictability surrounding the impacts of the climate change, the analysis of the climate and affected crops data will be an important tool for developing the targeted strategies to help farmers adapt to climate conditions and reduce the negative impacts of weather conditions.

The weather information provided by media is not adequate because it doesn’t give long-term expectations that can be used to help the farmers take the necessary precautions. Therefore, the main goal of this thesis is to propose and develop a new idea of predicting the agricultural risks by using spatial data mining theories and applying them to the data obtained from the Palestinian Meteorological Authority and the Ministry of Agriculture.

3. Motivation

A set of motivations stand behind analyzing the climate variability, including the financial benefits. The main objective of this research is predicting the agricultural risks for farmers due to the abnormal weather conditions. There are many weather websites and TV channels that broadcast weather on daily basis. However, those sources do not give long-term expectations that can be used to help the farmers take necessary precautions to reduce the potential losses. In addition, the details about the percentage of the affected crops would be invaluable for reducing the agricultural losses.

At the national and international levels, climate and agricultural development policies have a strong impact on poverty, the ways of living, food and human security. Better understanding of these impacts will lead to better results that will have a significant impact on human well-being and environmental sustainability.

This research will predict the agricultural risks based on data collected from the meteorological sites according to the proposed model in this research. These predictions will help farmers take precautions and safety measures to protect their farms from damage.

4. Contribution

Frequent weather phenomena are spreading beyond human control. However, it is possible to adapt to adverse weather conditions if weather prediction can be obtained in a timely manner.

Thinking about the food that we eat on daily basis, we have to think about the importance of the success of agriculture that could be achieved by protecting farms from the natural disasters. Many farmers lose their farms because of some diseases related to unexpected weather changes.

Agriculture is the most important field of research , especially in the developing countries, such as Palestine. Thus, the use of information technology and data mining techniques in agriculture can change the decision-making situation and farmers can achieve better results in their production.

Data mining plays an important role in predicting many agricultural issues. Therefore, the various data extraction applications are also discussed to address the various agricultural problems.

The contribution of our research is to predict the agricultural risks based on weather data collected from meteorological sites.

We applied spatial data mining methods to predict weather conditions which can be used to alert farmers to take precautions and avoid agricultural damage. Then we suggested a model that also applies spatial data mining methods using geographic and spatial data; however, it considers spatial parameters as variables and the effect of spatial topologic relationship between adjacent areas. The focus was to apply data mining theories in a new spatial form.

5. Background and Literature Review

5.1 Geographic Information System (GIS)

GIS can be described as an information system which is able to input, store, manipulate, analyze and output geographically referenced data. It is used to help taking decisions with regard to planning and management of transportation, land use, environment, and health services, etc. Other GIS tools include ArcGIS Desktop , GRASS, which stands for Geographic Resources Analysis Support System) and QGIS.[6]

ArcGIS is a set of tools that is used to manipulate spatial data and provide useful information which includes, but not limited to, inputting, storing and manipulating data, and reporting results [7].

Geometric location and information attributes concerning geographical features are stored using a non-topological format called a shapefile. The features can be expressed by lines, points and polygons. Shapefiles may also be associated with dBase tables for storing other information attributes that can be linked to the features of a shapefile. [8]

5.2 Data mining

Data mining is related to processing and analyzing data patterns based on a variety of categorization perspectives with the goal of concluding meaningful information. To enhance efficiency, data is gathered and classified into common areas. After that, data mining algorithms are applied to data for the purpose of having useful information that can be used to take decisions.

Data mining can also be thought of as a way to extract knowledge. Such a process is not easy as it involves large amounts of spatial data that is gathered for several applications like GIS remote sensing, environmental planning, etc [9]. Such data has

greatly enhanced human capacity for analysis and required improving database technologies and ways of gathering huge amounts of data like remote sensing and telemetry. This increasingly collected data required extracting knowledge of information from the data, which pushes toward the emergence of a new field, called data extraction and discovery in databases.

Although there are many studies of data extraction in relational databases and transactions, data mining is more demanding in other application databases, including spatial databases, time databases, objectively oriented databases, multimedia databases, and so forth [10].

The extracted knowledge may help to predict climate and understand its variability. Acquiring knowledge requires accomplishing several steps including selection, cleaning enrichment, transformation of data. It also requires data mining, reporting and displaying the discovered knowledge [11].

The selection of data requires identifying the appropriate source and type of data, as well as appropriate data collection tools. The actual practice of data collection is very important in its selection. Moreover, the process of selecting suitable data for a research project can influence data integrity. In addition, data selection should be carried out taking research questions into full consideration. The identification of data and its sources is often specific in a given area and is based primarily on the nature of the investigation, existing science and access to data sources.

Data cleansing or data cleaning refers to detecting and removing of corrupt or wrong pieces of data from a database table where incomplete or incorrect pieces of data are identified. The process may also involve fixing incorrect data. Moreover, Data is interactively implemented or cleared with data interleaving tools, or batch processing through scripting [12].

Data enrichment refers to enhancing raw data. This includes any process with the goal of making data more valuable for the project. The enrichment of data shows the common need to use this data in different ways and in advance.

Data transformation involves modifying the format of data values from the source format to another format that is suitable for the destination data system. Data conversion usually involves two phases:

1. Mapping data elements of the source system to the destination system, where the possible methods are determined to make the transformation.

2. Generating the required code to make the transformation.

It's very important to use data mining in agriculture because this field contains a lot of data, such as soil data, crop data, weather data, and so on. Several data mining algorithms are used to analyze the agricultural data and provide useful pattern. K-Means clustering, Apriori algorithms and other statistical methods are examples of such algorithms. Moreover, data extraction software is an analytical tool used for analyzing data using different perspectives. The process includes classifying data and discovering possible relationships. It may also include clustering and regression. [13]

Data mining is now used in various fields, including time series data which is used for weather prediction with the help of data extraction techniques. For time series data analysis, intelligent prediction models perform better than those traditionally used in forecasting. Neural network (n) and genetic algorithm (Ga) are two of the most popular techniques based on arithmetic intelligence.[14]

The graphical representation of different data mining techniques is shown in

figure 1 below.

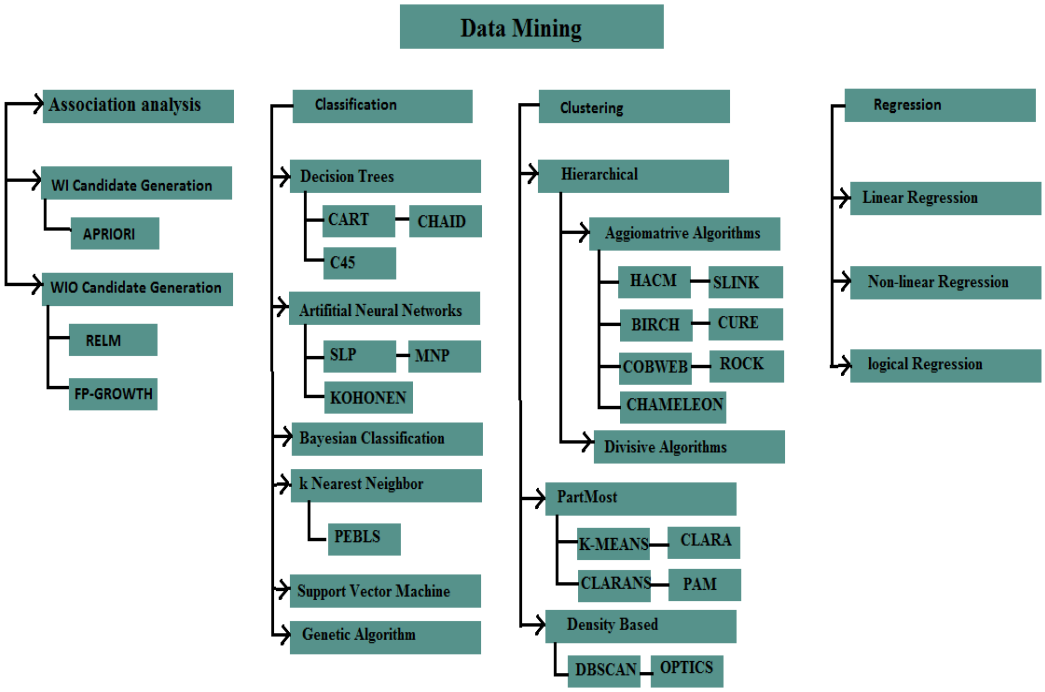


Figure 1- Data Mining Techniques[15].

The main techniques for data mining include association rules, classification, clustering and regression. Association rule mining technique is one of the most efficient techniques of data mining to search unseen or desired pattern among the vast amount of data. Classification and prediction are two forms of data analysis that can be used to extract models describing important data classes or to predict future data trends. In clustering, the focus is on finding a partition of data records into clusters in such a way that the points within each cluster are close to one another. Also, regression is learning a function that maps a data item to a real-valued prediction variable.

5.3 Neural network

The term ‘Neural’ is related to the “neuron”, to the nerve cell of the human that exists in the brain. After the conclusion that the human brain works in a completely different

way from the traditional digital computer [38], the use of the artificial neural network has become a catalyst [33][34]. The brain has the ability to reconstruct its neurons in a way that enables it to perform complex calculations very quickly compared to the computer. However, the brain performs many tasks , such as sensory perception and face recognition, which are complex tasks that could take days on a traditional computer.

5.3.1 Structure of Neurons in Brain

Figure 2 shows the main components of a biological Neuron:

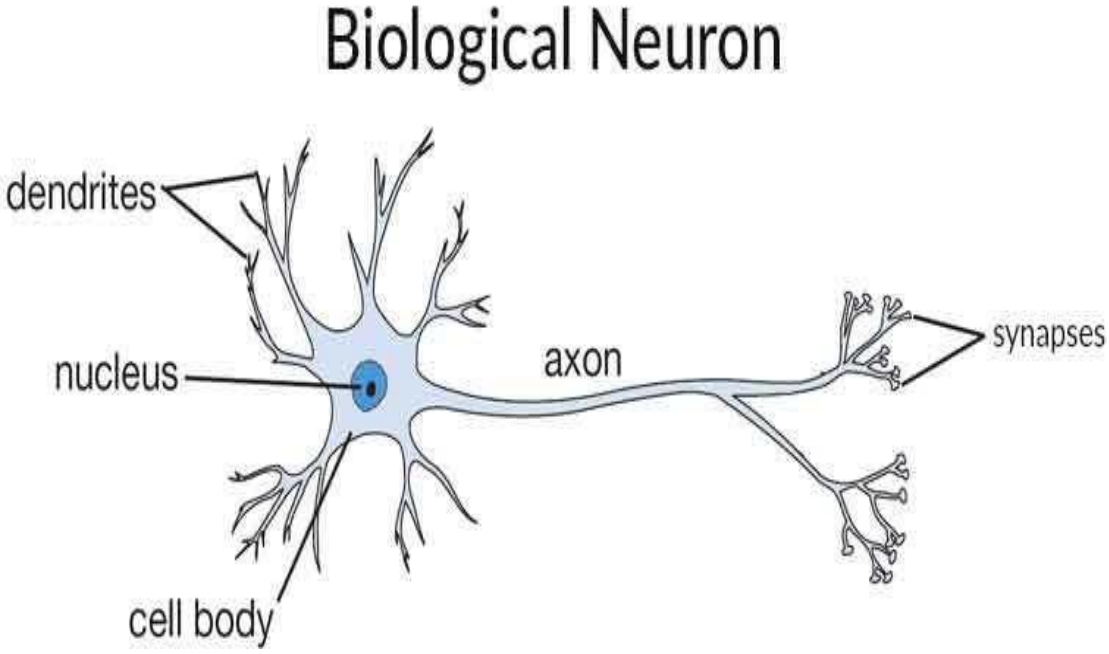


Figure 2- The typical nerve cell of human brain comprises of four parts[16]

Dendrite— gets signals from other neurons.

(cell body)—gathers all incoming signals to generate input.

Axon—used to allow the signal to travel down to the other neurons.

Synapses—The point of interconnection of one neuron with other neurons.

5.3.2 Advanced Neural Network (ANN) architectures

There are many different Advanced Neural Network (ANN) architectures which can be used for the prediction of data. The most common techniques are mentioned below:

1. **Single Layer Feed-Forward Network**

   In a single layer neural network, source nodes are organized in an input layer and output layer. These nodes make the input for an output layer of neurons. This kind of organization is referred to as a feed forward type. The architecture is shown in figure3 below.

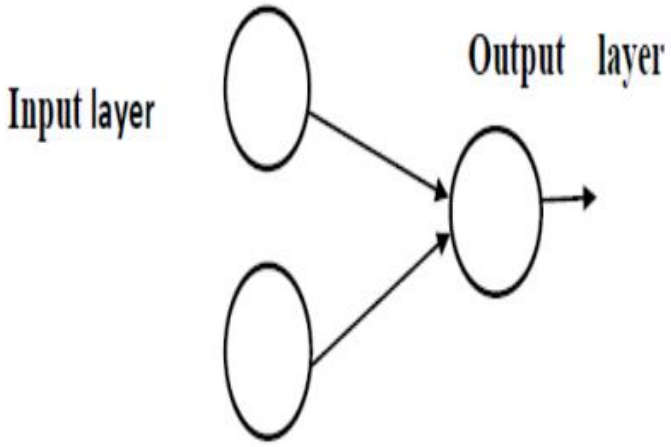


Figure 3- Single Layer Feed Forward Network[17]

2. **Back-propagation Network (BPN)**

   BPN is also a feed-forward network. It has three layers; input, hidden and output layers. The hidden layer can be comprised of more than one layer based on the problem complexity. Less number of layers reduce computational time

and complexity of training. The architecture of this network is shown in figure 4 below.

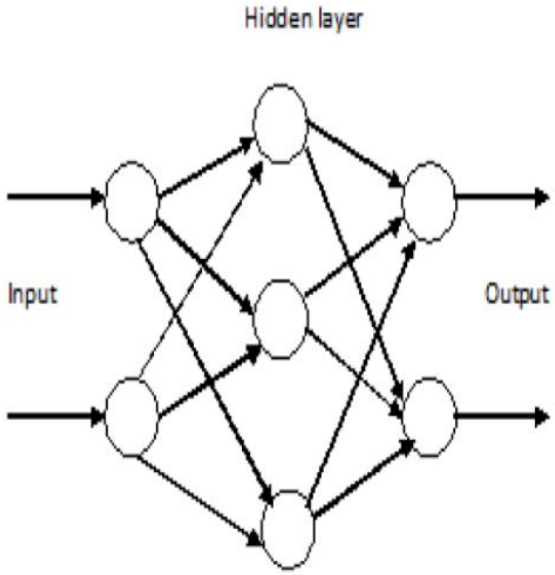


Figure 4- Architecture of Neural Network with Hidden Layers[17]

The number of layers and the number of other elements are made by the programmer while building, training and testing the network, which are very important decisions. The process of training involves applying input data to be the input layer and comparing the data with the output layer with the desired output. The difference is forwarded back to the previous layers. An algorithm called gradient descent algorithm is applied to minimize the mean square error between the output layer of the network and the desired output. The performance of a neural network is affected by the chosen weights and the input-output function specified for the units [17].

3. Radial Basis Function Network (RBF)

RBF network has three layers: an input layer, an output layer and a hidden layer. A feature that distinguishes this network is that each hidden unit in a hidden layer implements a radial activated function. This makes this type of network more accurate and faster than feed-forward networks. Similar to the network in the previous section, the gradient descent algorithm is used to minimize the error between the target and the desired output [18]. The architecture of the network is shown in figure 5.

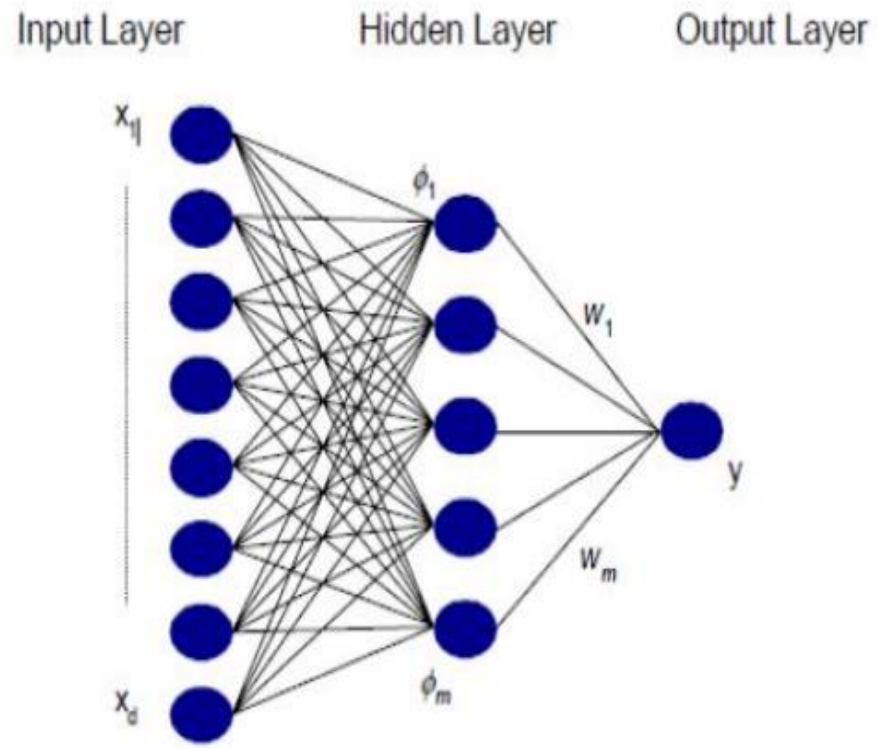


Figure 5- RBF Network Architecture[19]

4. **Adaptive neuro-fuzzy inference system (ANFIS)**

   ANN and Fuzzy logic are used in an important and effective technique in engineering problems. Fuzzy logic uses rule-based modeling human thinking and decision-making. On the other hand, ANN learns the problem through successfully for data sets and by using its ability of learning .Jang [19] suggested the method of ANFIS considering the fuzzy logic methods and advantages of ANN. ANFIS has come from the integration of ANN and fuzzy inference systems.

5.3.3 The most Common Types of Learning in Neural Network

Supervised Learning: Inputs are training data for the network. Moreover, weights are adjusted until the desired value is obtained.

Unsupervised Learning—Input data are used to train the network whose output is known. The network adjusts weight by extracting a feature in the input data.

Reinforcement Learning—It is semi-supervised learning where the output is unknown. Feedback is always generated and it is not dependent on the accuracy of the output.

5.3.4 Learning Data Sets in ANN

Training set: A set of inputs and desired outputs that are used for learning. It is used to choose suitable parameters of the network

Validation set: A set of inputs and outputs that help choosing the architectural parameters, like the number of hidden units of the network.

Test set: A set of inputs and outputs that are used to test the performance of a network.

The neural network is said to be learned by updating the weights inside the network after several iterations.

5.3.5 Four Different Uses of Neural Networks

Classification—These kinds of networks are feed forward networks that are trained to be able to classify data set into predefined class.

Prediction—These networks are used to predict outputs based on the given inputs. Networks used for weather prediction are examples of such networks.

Clustering—These networks are used to categorize data and identify a special feature of it although they have no idea about the input data.

The following networks are used for clustering –

- Competitive networks
- Adaptive Resonance Theory Networks
- Kohonen Self-Organizing Maps.

Association—This kind of network is used to remember a certain pattern, so that when the noise pattern is presented to the network, the network will be able to associate it with a similar one in its memory or discard it [38].

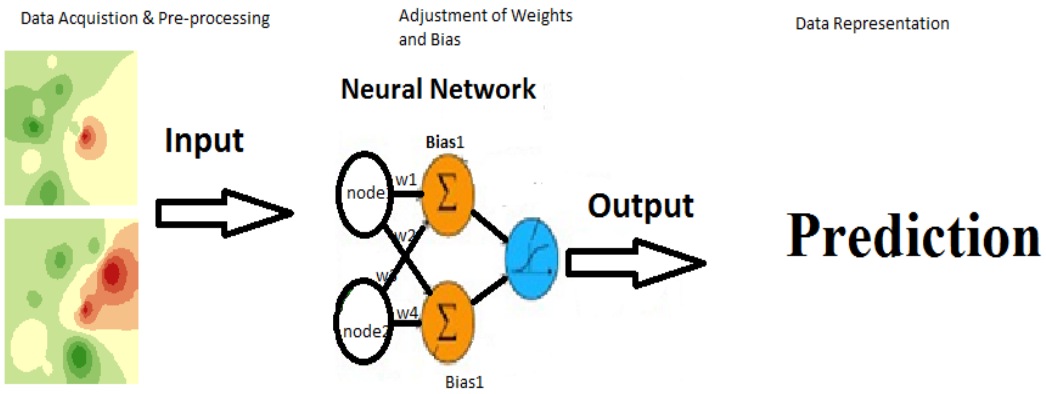


Figure 6- Neural Network for Pattern Recognition

A well-known example of association is called pattern recognition. It refers to the process by which machines can notice the environment, learn patterns of interest, and

make decisions regarding the category of the pattern. Figure 6 shows a neural network for pattern recognition. Examples of pattern recognition networks are those used to recognize fingerprints and speech.

5.4 Normalization

Normalization is normally done, when there is a distance computation involved in our algorithm, like the computation of the Minkowski dimension.[20]

Some of the techniques of normalization are:

1. Min-Max Normalization – In this technique, data is fit in a pre-defined interval[C,D].

   Formula -

$$B = \left(\frac{(A - \text{minimum value of } A)}{(\text{maximum value of } A - \text{minimum value of } A)}\right) * (D - C) + C$$

2. Decimal Scaling - In this technique, the computation is generally scaled in terms of decimals. It means that the result is generally scaled by multiplying or dividing it with pow(10,k).
3. Standard Deviation method - In this method, the d is normalized by using the formula [x-mean(x)]*sd(x)
4. By eliminating outliers - Outliers are a common sighting while dealing with data. Their presence creates quite a lot of hassles in the computations. So, eliminating them is a very clever idea. So, detect your outliers from the box-plots and refine your data by eliminating them.[21]

Data preprocessing is very important. It refers to the process of extracting, cleaning, and transforming to a format, where data mining algorithms could be applied. A necessary step is called normalization, where parameters of different units and scales are unified.

One possible formula for rescaling is given below where min and max are the minimum and maximum values in X, where X is the set of observed values of x [22].

$$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

In the data preprocessing step, the data in GIS format is extracted, cleaned, and transformed to a format, where data mining algorithms could be applied to the data mining step.

5.5 Interpolation

Environmental data are collected worldwide using specific stations that are distributed over a geographical area. Collected data is used for the researchers to help them plan well and make appropriate decisions. However, data collected by the stations doesn't cover all locations; however, it is needed by the researchers to be able to accomplish their research. To solve this problem, interpolation techniques are to estimate missing data based on mathematical functions and methods [23]

There are different interpolation methods that can be used based on the field of research. For the environmental and geographic information, the scientific research involves spatial data since interpolation is based on the assumption that attribute data is continuous above the space. This allows attribute estimation anywhere within data boundaries. Another assumption is that the spatial attribute is dependent on the closer values. These assumptions allow spatial interpolation of methods to be formulated. Some of the interpolation techniques include Kriging, IDW (Inverse distance weighted), Splines, and TSA (Trend surface analysis). They can be used to estimate

the lost environmental data, such as temperature, precipitation, and toxic substance concentrations [24].

The huge data volumes available from GIS make it realistic to include additional information into the processor estimate. However, traditional interpolation rarely explicitly includes many approaches to this information available. This observation suggests existing knowledge technologies that can lead to new better methods of interpolation [25].

5.6 Space Time Cube

Space time cube presents a set of points into a netCDF data structure. NetCDF (network Common Data Form) is a file format for storing multidimensional scientific data (variables) such as temperature, humidity, pressure, wind speed, and direction. It shows the points using space-time bins. Within each bin, the points are counted. The counts over time are counted for all bins locations.

A time series represents a group of values measured sequentially over time. Data mining of a time series results from the goal of viewing the shape of data. The cube is made of rows, columns, and time steps. The total number of bins can be calculated by multiplying the number of rows by the number of columns by the number of time steps. The rows and columns are used to represent the spatial extent of the cube, while the time steps represent the temporal extent. Locations with data are places (bins) where at least one point has occurred over time [8].

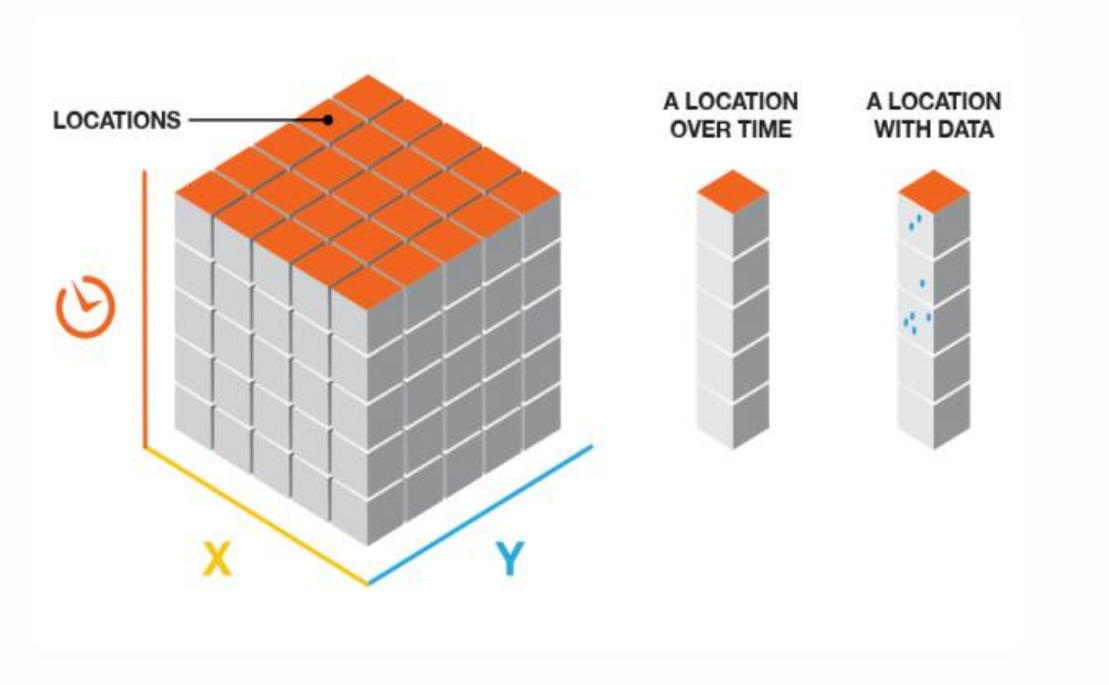


Figure 7- Space time cube (A time series as space-time bins). [8]

Usually, locations with data for at least one time-step interval will participate in the analysis, but they will be analyzed across all time steps.

5.7 Cross - validation and multilayer perceptron network

Cross-validation represents an alternative to the residual evaluation models. It is considered better than residual evaluations. especially in situations where predictions are required for data that has never been seen. Cross validation solves this problem by using part of the data when training a learner. They exclude some of the data before training begins and use it to test the performance of the learned model on new data [26].

One of the simplest types of cross validation is called the holdout method. In this method, data set is separated into two sets called the training set and the testing set. A function called approximate or fits a function using the training set only. Then the function approximation is asked to predict the output values for the data in the testing

set although it has never seen these output values before. The errors are summed and used to evaluate the model. This method is relatively fast (compared to the residual method). However, the evaluation may be highly dependent on data points that are chosen to be in the training set and those on the test set.

To reduce the dependency on the division between training set and testing set, another method called K-fold cross validation is used. The main idea is to divide the data set into k subsets, and the holdout method is repeated k times taking one of the subsets as test set and the other k-1 subsets as a training set. The average error is computed and the variance of the resulting estimate is reduced. A disadvantage of this method is that it is relatively slow since it has to compute the holdout method k times. A variant of this method is based on choosing test and training sets randomly [27].

Leave-one-out cross validation is K-fold cross validation, where K is equal to the number of data points in the set. As before. the average error is computed and used to evaluate the model. The evaluation given by leave-one-out cross validation error (LOO-XVE) is good, but the method is relatively very slow. Fortunately, locally weighted learners can make LOO predictions just as easily as they make regular predictions.

A multilayer perceptron is made of a number of layers; each has one or more neurons. where input neurons (input layer) feed input patterns into the rest of the network. The layers after the input layer are called hidden layer and they are followed by a final output layer which holds the results. Each unit in a layer is connected to all units in the subsequent layer and each unit receives an input from all units in the preceding layer. Moreover, each connection has a certain weight which indicates the strength of the unit relative to the unit in the subsequent layer. This arrangement makes the output dependent on the input and on the connections weights of the units. When

information is presented to a multilayer perception, this information propagates layer by layer until finally the output layer is activated. It has been shown that multilayer perceptron can virtually approximate any function with any desired accuracy. However, this requires having enough hidden layers and sufficient training data. In other words, accuracy is highly dependent of the amount of training data [28].

Small training set can lead to the case when a network failed to distinguish between pattern data and noise. To avoid over fitting, a feature called generalization is used. It is mainly based on using 80% of the data as a training set and testing the trained network with 20% of the data.

5.8 Weka

Weka is a machine learning tool that can be used to implement the neural network. It is built using JAVA programming language which makes it an open source tool. Weka was developed at the University of Waikato, New Zealand. It has a graphical user interface ,which makes it easy to run a regression or classification scenario. However, experienced JAVA programmers may decide to use a command shell or build their own classes.

Weka has several learning modules , such as linear regression, neural networks (a.k.a multilayer perceptrons), decision trees, support vector machines, and even genetic algorithms [29]. Weka uses a format called ARFF format.

5.9 Related Work

Climate change leads to the emergence of dynamics and uncertainties about agricultural production. There are many models that have been used to show the possible consequences of the climate change based on different climate change scenarios, such as Combinations of General Circulation Models, soil models, agro-ecological system models, Regional Circulation Models and economic models [30], [31], [32] and [33].

There are many factors affecting the weather, which are directly related to animal and vegetable farms, such as humidity, rains, temperature and wind speed. The following authors analyzed geographical and climatic data and their impact on crop growth and agricultural and animal damage caused by inappropriate climate conditions using data extraction techniques.

Swati Hira et al (2015) built a multidimensional model of data and then applied a multidimensional analysis. Consequently, they came up with that agricultural data are temporal spatial data, which include agricultural parameter data, environmental features and geographical characteristics. These data must be analyzed through a multidimensional analysis, statistical analysis and data extraction techniques (AMS) to obtain a useful pattern, which helps to analyze agricultural productivity. A multidimensional model was built before the multidimensional analysis. IDASM is a tool used to build a multi-dimensional model and perform statistical techniques and extract data, which provide the relationship between different agricultural parameters. They concluded that the agricultural data grew exponentially, and the use of data extraction techniques or traditional tools, such as spreadsheets, are not strong enough to extract data patterns since it has failed to produce strategic information. The authors used IDASM tool to build multidimensional model to find the relationship between the parameters of agriculture. Moreover, they used data mining techniques and applied statistical mining to get better results in finding parameter relationships. [34]

Harln D. Shannon (2015) examined the various natural disasters resulting from weather and climate that occurred in the agricultural land in North America and Central America. As recent history of climate and weather data helps farmers manage agricultural risks. The research discussed climate risks in agriculture such as drought,

floods, hurricanes, extreme heat and freezing. The decision support system was used for farmers to take all precautionary precautions before the disaster.[35]

Laila Mohamed and her colleagues used data mining techniques and show the number of animals where affected for each disease when the humidity is low and temperature is hot. They developed a system using data mining technique to analyze and measure the climate effects on the animal production. The developed system combines the BOVIS database with the weather database through data warehouses techniques. The system results of analyzing discovered knowledge showed that their system can be used for prediction of time occurrence of the disease. They suggest to use more attributes and increase them from attributes which are not used in both BOVIS and weather database such as , animal feed and rain. Also they propose to integrate their system with other data mining algorithms to provide more discovered patterns [31]. From the last years meteorological databases have enough data in the farms production that will give chances to researchers to apply and model new algorithms for farms losses due to the weather.

Vale and his colleagues analyzed poultry production databases related to the climate data using the data mining techniques; attribute selection, normalization, data classification and decision trees were used for the prediction of the effect of heat waves on broiler mortality. The authors proposed models, which depend on the data collected from the meteorological stations that produce huge amounts of data that are rarely used for animal production or agriculture. Consequently, the authors suggested that the development of the models for prediction of the production losses for animal and agriculture could be useful for farmers. [32].

The meteorological data mining is a form of data mining that is concerned with finding hidden patterns for the available meteorological data, so that it can be

retrieved and transformed into usable knowledge. Sara N and her colleague used data weather collected locally in Gaza city to extract useful knowledge. The data included nine years period from 1977 to 1985. The authors applied several algorithms , such as analysis, clustering, prediction, classification and association rules mining techniques, then they presented the extracted useful patterns and described the importance of it in the meteorological field. They proposed building new adaptive model and dynamic data mining methods that can learn the nature quickly and dynamically for weather conditions changes and sudden events as a future work. [45].To make data mining algorithms more accurate in predicting, it's important to store the meteorological data where it becomes cumulative over time; in addition to that it's important for taking the new data, which can occur suddenly.

Disasters Saptarsi Goswami and his colleague discussed the application of data mining and analysis methods for prediction, detection and development of disaster management which depends on data collected from disasters centers. They collected data from the available recourses , such as, satellites, remote sensing and newer sources like social networking sites. The authors proposed a framework for building a disaster management database to predict disasters in India Big Database platform like Hadoop. They were concerned that there was not enough work done in this area to take advantage of the data sources and possibilities of their availability, especially in the country of India, so they propose to store a natural disaster data as a first stage and integrate this information with other sources of information as another stage. [36].

Also Folorunsho Olaiya and his colleague checked the using of data mining techniques in prediction the maximum of temperature, rainfall, evaporation and wind speed. Where the authors show that their results that given from large data over time could be analyzed and show deviations which show changes in climatic patterns

discovery. The authors used Artificial Neural Network and Decision Tree algorithms for data collected from meteorological stations from 2000 to 2009 from Ibadan city in Nigeria. the authors proposed to use neuro-fuzzy models as the future work for the weather prediction process. This work is important to climatic change studies because the variation in climate conditions for the attributes of temperature, rainfall and wind speed. [37].

Mallari et al (2015) Predicted that vulnerability assessment was a useful means of increasing agricultural sector adaptation to the climate change, the vulnerability assessment is a method that can improve farmers' decision-making, which may increase the resilience of agriculture systems during the hazard events. The city of Mapalact has been considered for this research to assess vulnerability using the following methods: 1.Index method and 2.Geographic information systems (GIS).

In the index method, three types of vulnerability indicators were selected, such as sensitivity indicators, vulnerability indicators and adaptive capacity indicators. These indicators help GIS to predict a location that is highly vulnerable to climate change. Finally, a map was created that enabled farmers to reach the best agricultural pattern. [44]. The table1 below shows the summery of the related works:

Reference #	**Proposed**	**Finding**	**limitation**
Laila M, Maryam H, Alaa Eldin Y. 2012	a system using data mining technique to analyze and measure the climate effects on the animal production	They showed that their system can be used for prediction of time occurrence of the disease.	use more attributes and increase them from attributes which are not used
Vale, M. M., Moura, D. J., Naas, I. de A., Oliveira, S. R. de M., and Rodrigues, 2008	represents a good base for analysis and predictions in the following time period for the purpose of quality decision-making	improve animal production by studying the impact of climate change on animal production.	using more attributes from the unused variables also integrate it with others data mining algorithm
Sarah N. Kohail, Alaa M. El-Halees 2010	build new adaptive model and dynamic data mining methods that can learn the nature quickly	obtain useful prediction and support the decision making for different sectors.	building adaptive and dynamic data mining methods that can learn dynamically to match the nature

DisastersSaptarsiGoswami 2016	propose framework for building a disaster management database to predict disasters	a framework for building a disaster management database hosted on open source Big Data platform has been proposed.	there have not been enough works done to tap the potential of these sources especially in context
Swati Hira et al 2015	built a multidimensional model of data then apply multidimensional analysis,	they used data mining techniques and applied statistical mining to get better results in finding parameter relationships.	use more attributes and applying other models for the weather prediction process. And make comparisons.
Harln D. Shannon 2015	examines the various natural disasters resulting from weather and climate	The decision support system was used for farmers to take all precautionary precautions before the disaster	The rsearchers didn't cover all variables that affect on agriculture.
FolorunshoOlaiya 2012	classifying weather parameters such as maximum temperature, minimum temperature, rainfall, evaporation and wind speed	show that given enough case data, Data Mining techniques can be used for weather forecasting and climate change studies.	use other models for the weather prediction process.
Mallari et al 2015	assess the vulnerability of the agriculture sector	vulnerability assessment used in generating planning measures which can increase the resilience agriculture sector to the impacts of climate change	need for the local government unit to generate measures to reduce the vulnerability of agriculture sector to climate change.

Table1: shows the summery of the related work

Our suggested work will take advantage of other related works, so we propose to collect data for all attributes from metrological stations and the Ministry of Agriculture that will help us in forecasting the weather with the agriculture risk at the same time.

The data will be prepared to analysis in excel sheet and some operations will be done for the data , such as normalization and interpolation before using analysis mechanisms, then data mining methods will be used to predict any sudden weather changes.

6. Experimental Setup and Work

The main objective of this research is to predict the agricultural risks based on weather conditions. The main steps/methods to achieve this goal can be summarized in the following points:

1. Data collection and preparation.
2. Georeferencing Map and Join data to the ArcGIS map.
3. Data interpolations
4. Applying several algorithm for percentage of affected crops prediction
5. Suggested Improvement Method of spatial analysis and data mining

6.1 Data collection and preparation.

In this step, weather data attributes were collected (wind velocity, humidity, rainfall, temperature, sunrise, atmospheric pressure) that help for forecasting. In additions, the collected data included the agricultural damage in terms of the number of people affected, damage rate and number of affected species in Jenin area on a monthly basis (2011, 2012, 2013, 2014 and 2015). The data used in the analysis was obtained from the Palestinian Meteorological Authority and the Ministry of Agriculture. The study includes eleven sites including Jenin city and its surrounding villages. These sites were chosen because they represent one of the most suitable areas for agricultural production in the West Bank[42]. Figure 8 shows the study area.

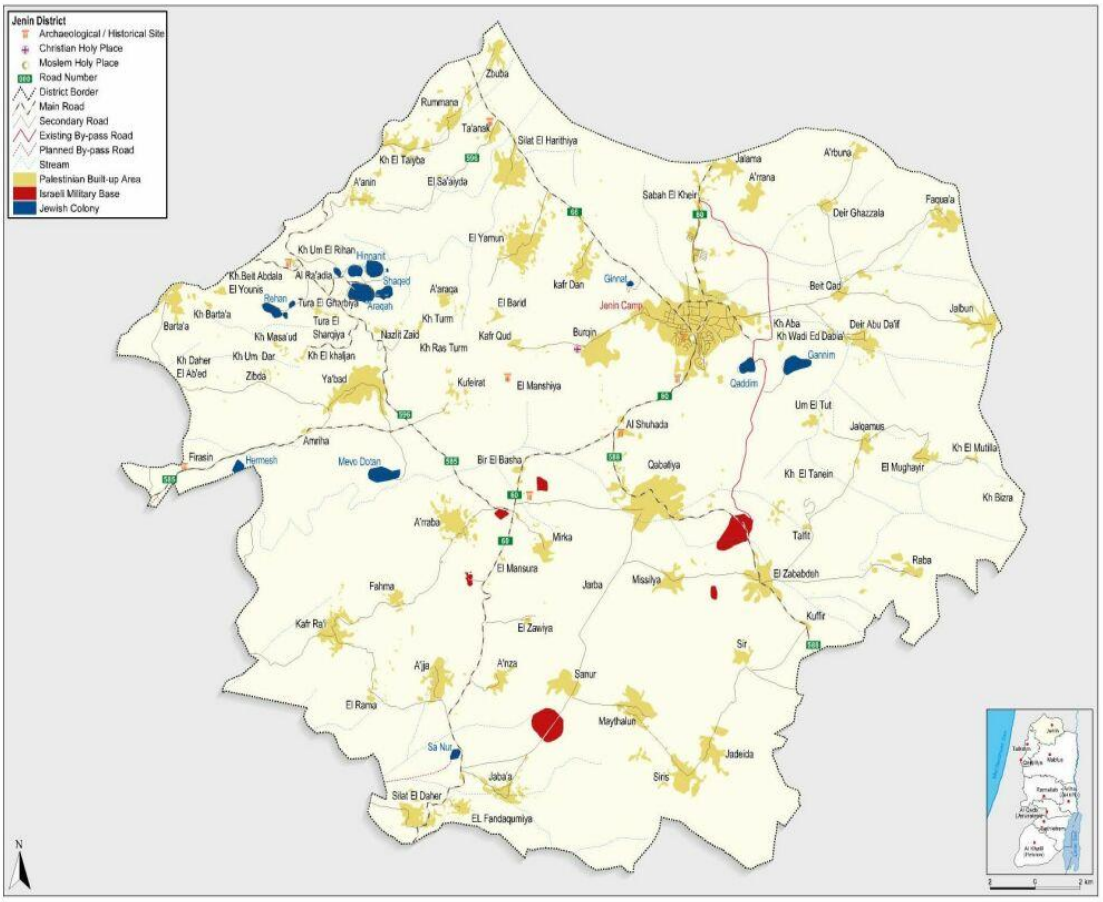
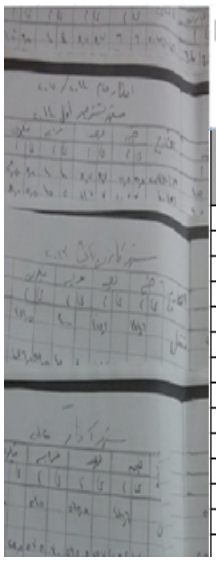


Figure 8- Jenin district our study area[39]

Figure 9 shows the sites involved in the study and the format of the collected data. It shows part of the attributes , where each attribute occupies a column in the spread sheet



rain p4	rain p3	rain P2	rain p1	date	rain	tempav	RH	sunshine	evap	wind	rofAffecte	AffectedP ercentage	Numberof PeopleAff ected	area
0	0	5.810723	5.858444	Jan-12	5.824996	4.46	9.94	4.34	0.40	7.08	0.00	0.00	0.00	1
0	0	5.842647	5.898167	Jan-12	5.858444	4.46	9.94	4.34	0.40	7.08	0.00	0.00	0.00	2
0	0	5.880687	5.945928	Jan-12	5.898167	4.46	9.94	4.34	0.40	7.08	0.00	0.00	0.00	3
0	0	5.926703	6.004043	Jan-12	5.945928	4.46	9.94	4.34	0.40	7.08	0.00	0.00	0.00	4
0	0	5.983232	6.07547	Jan-12	6.004043	4.46	9.94	4.34	0.40	7.08	0.00	0.00	0.00	5
0	0	6.053648	6.163785	Jan-12	6.07547	4.46	9.94	4.34	0.40	7.08	0.00	0.00	0.00	6
0	0	6.142210	6.272834	Jan-12	6.163785	4.46	9.94	4.34	0.40	7.08	0.00	0.00	0.00	7
0	0	6.253791	6.405558	Jan-12	6.272834	4.46	9.94	4.34	0.40	7.08	0.00	0.00	0.00	8
0	0	6.392804	6.560604	Jan-12	6.405558	4.46	9.94	4.34	0.40	7.08	0.00	0.00	0.00	9
0	0	6.560016	6.723963	Jan-12	6.560604	4.46	9.94	4.34	0.40	7.08	0.00	0.00	0.00	10
0	0	6.743731	6.854916	Jan-12	6.723963	4.46	9.94	4.34	0.40	7.08	0.00	0.00	0.00	11
0	0	6.900098	6.88736	Jan-12	6.854916	4.46	9.94	4.34	0.40	7.08	0.00	0.00	0.00	12
0	0	6.942963	6.78703	Jan-12	6.88736	4.46	9.94	4.34	0.40	7.08	0.00	0.00	0.00	13
0	0	6.824703	6.599189	Jan-12	6.78703	4.46	9.94	4.34	0.40	7.08	0.00	0.00	0.00	14

Figure 9- Collecting information and converting raw data to excel sheet

The next step after collecting data and storing it in the format explained in the previous paragraph is to prepare data for analysis using spatial data mining theories. As part of data preparation, data has been normalized so that values fall in the 1 – 10 scale for all attributes by dividing each value by the maximum value multiplying by 10. This step is important because it prepares data to be mapped to maps in the ArcGIS system for storing, presenting and analyzing data in later steps.

6.2 Georeferencing Map and Join data to the ArcGIS map.

In step 2, the data points shown in the scanned map in figure 10 are mapped to theArcGIS system map by a process called georeferencing. Several control points at the corners and middle of the map were used to assure accurate mapping [40].After that, names were given to the study points on the ArcGIS program using the same names that we used in the collected normalized data in excel sheet. Then, ArcGIS is used to generate a table that has one row for each study point. The ArcGIS table (data layer) is then filled with our collected data using the ArcGISJOIN function based on the names of the points that are the same in both the ArcGIS map and the excel file.

Now, as we see in figure 10 we have the study points in Jenin strict and their attributes (data) entered to the ArchGIS system so that we can start data mining as explained in later steps.

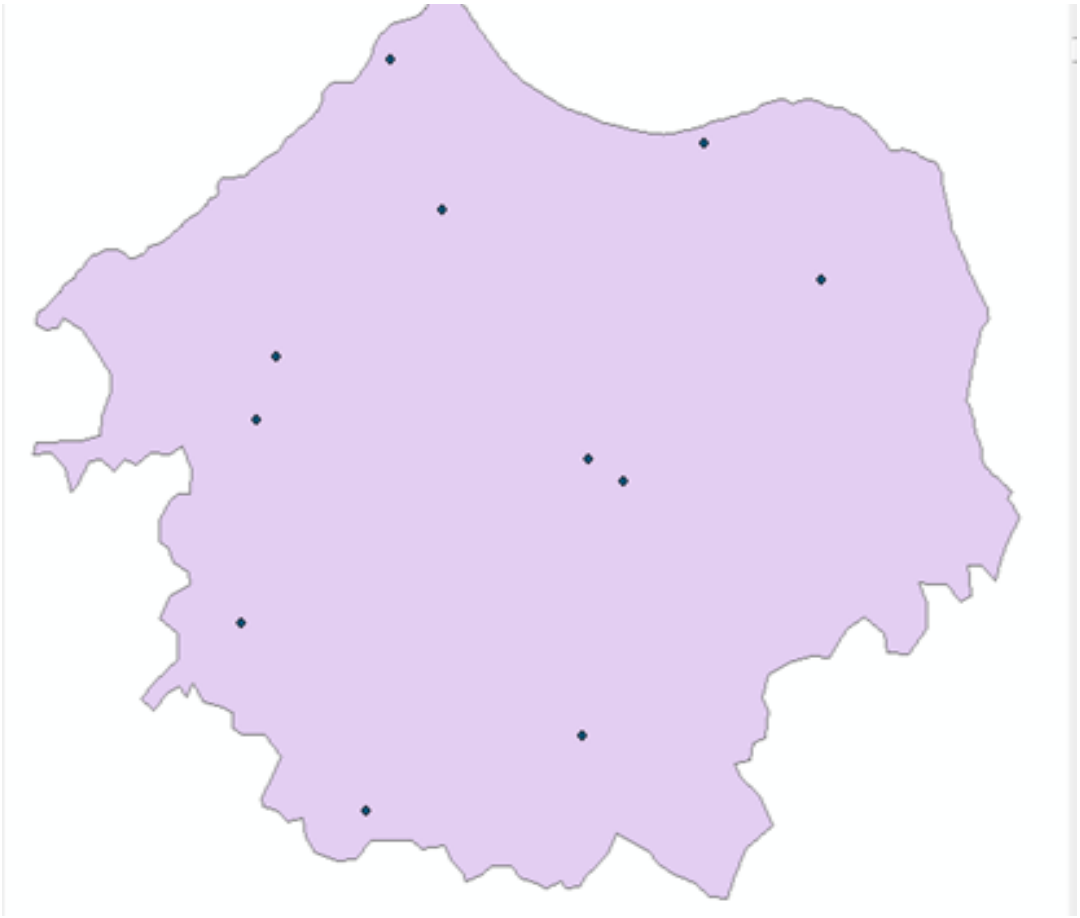


Figure 10 - join xls data file with areas in jenin area

6.3 Data interpolations

We used IDW interpolation in the current literature to satisfy the data environment. IDW is an averaging process; all interpolated values are within the sample range. It is commonly used for internal updating and it means estimating the environmental data for all the other points that do not not have data in the study area. The collected data represents monthly weather information about 11 points (areas) that are distrusted in

Jenin area. IDW interpolation created a raster image for the layer of Jenin map, which represented the estimated data for Jenin area, to be used to generate data for 850 points (25 rows x34 columns) based on the original 11 points.

We choose this resolution because the ArcGis read data from excel file (xls), as we know excel file doesn't support huge data (more than 65000 record) so 850 point for every month and 10200 record every year and 51000 records in five years. So 850 point for the raster is the maximum number that the excel file can deal with it among long time.

Each raster was converted into 850 (25x34) points in Jenin map and the estimated weather data were stored in new table as a matrix. This number came as a result of choosing a point for each 1000 pixels as shown in figure 11 below.

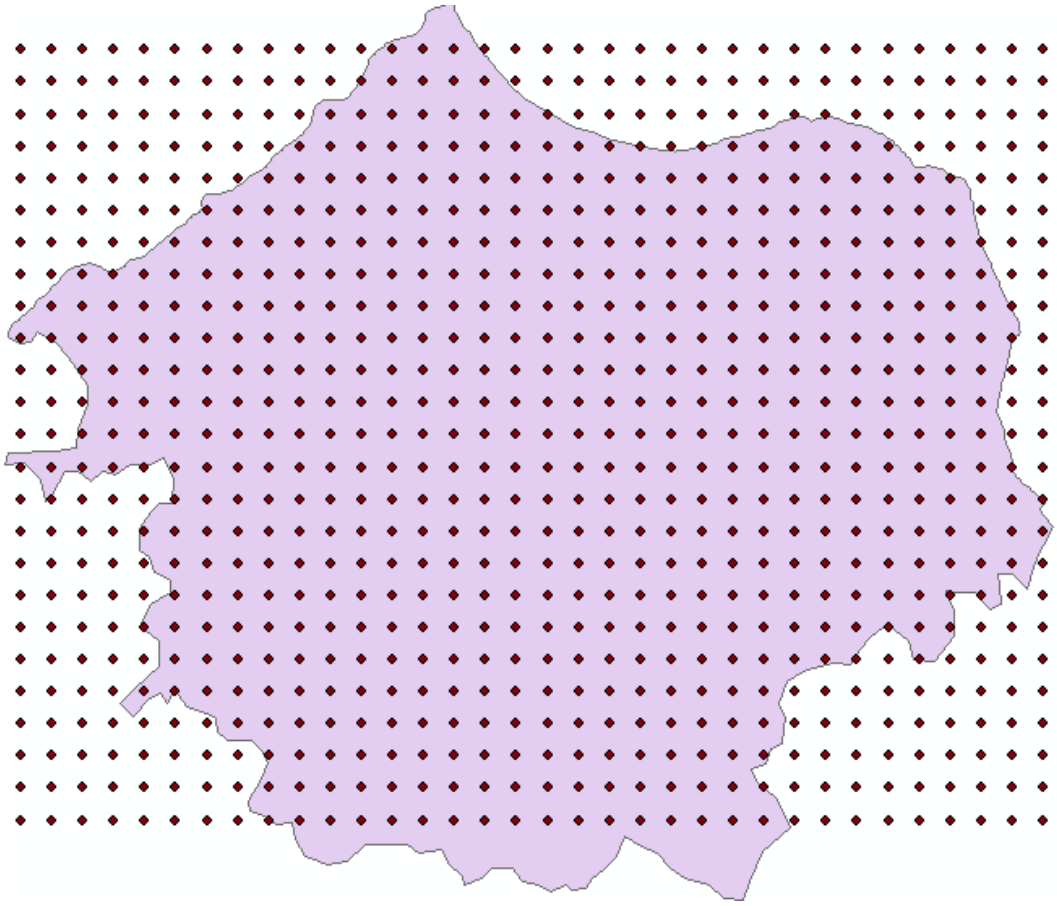


Figure 11- 850 new point for each attribute every month after interpolation.

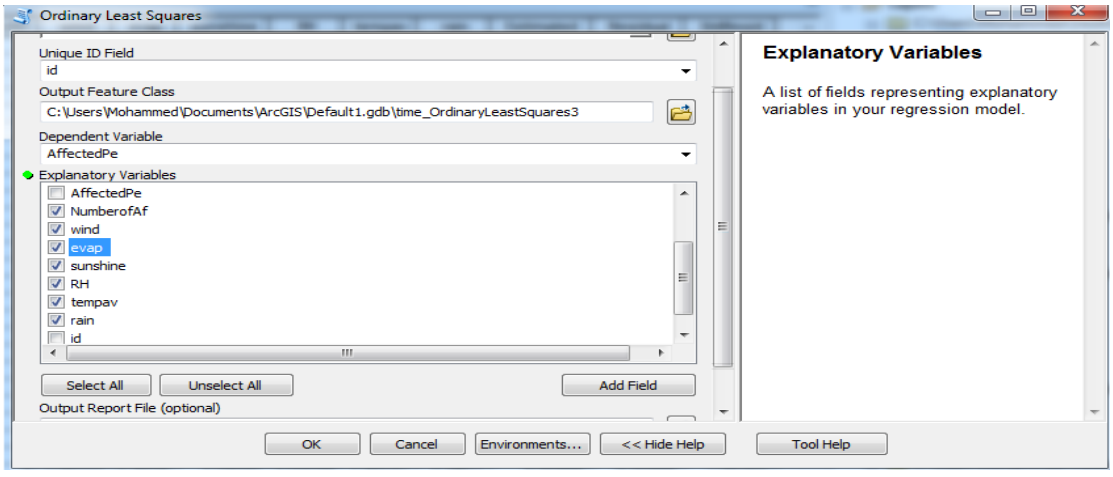
6.4Applying several algorithm for percentage of affected crops prediction

Several algorithms were applied to predict the percentage of affected crops, which are: Ordinary Least Square and multilayer perceptron neural network for our data after data interpolation. Also, the researcher proposed a new approach in prediction by taking the neighbors of each point and applying that approach in multilayer perceptron neural network.

6.4.1 Ordinary least square

The researcher applied Ordinary Least Square to the data and chose the affected percentage of agriculture as dependent variable and then used (wind, evaporation, sunshine, RH, temperature average, rain) as explanatory variables.

The Ordinary Linear Regression (OLS) resulted in the generation of a dependent variable prediction or model as it is related to a set of explanatory variables. After running the OLS tool, there were several inputs, which have to be filled with the appropriate variable: first, the Unique ID Field, second, the dependent variable, which is the variable to be predicted, third, the list of the explanatory variables. in addition to determining a path for the Output Feature Class. And there are several paths for the Output Report File, Coefficient Output Table, and Diagnostic Output Table as shown in figure 12.



Ordinary Least Squares Parameters

Parameter Name	Input Value
Input Features	time
Unique ID Field	id
Output Feature Class	None
Dependent Variable	AFFECTEDPE
Explanatory Variables	AREA
	NUMBEROFPE
	NUMBEROFAF
	WIND
	EVAP
	SUNSHINE
	RH
	TEMPAV
	RAIN

Figure 12- applying ordinary least square on our data

6.4.2 Applying multilayer perceptron neural network using Weka

In this section, the researcher used another tool, called Weka, to make prediction for agriculture risks. This tool uses neural network algorithms for prediction. Weks requires the data file to be stored as an arff file. However, there is no tool available for direct conversion from Excel files to arff file, so we built a java tool to convert excel data file to arff file.

Weka 3.8 is used to predict the agriculture affected percentage using cross validation and multilayer perceptron network for the data before and after interpolation to study the effect of weather conditions and other geographic parameters on agriculture based on the available meteorological and agricultural collected data.

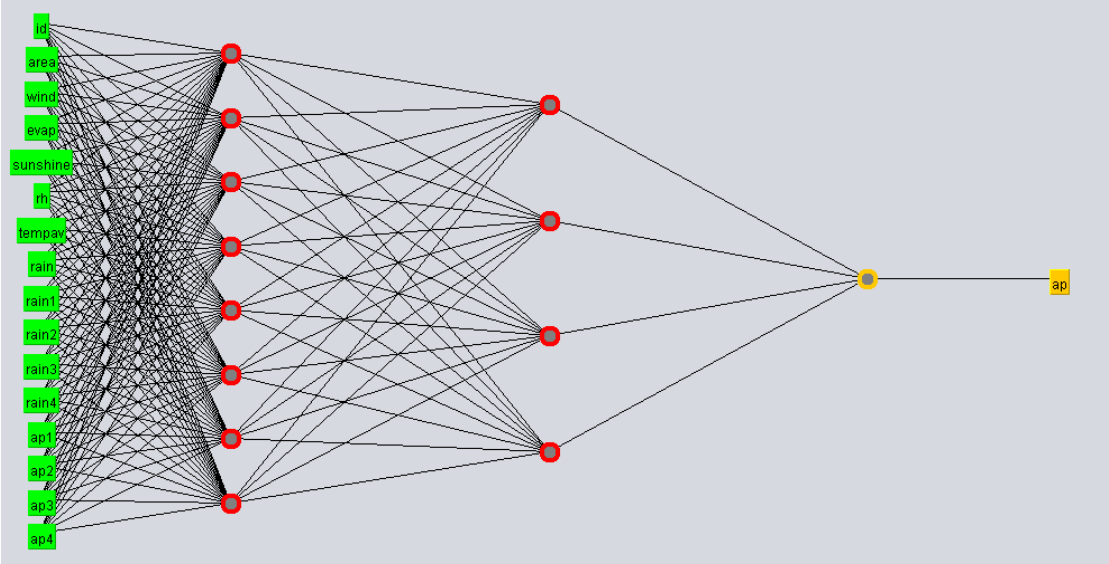


Figure 13 - our model in prediction in multi perceptron neural network using 2 hidden layer.

The model of neural network generated by Weka is shown in figure 13. Also, Weka provides results with a summery for the prediction process as a plain text displayed in figure 14 below, the actual , predicted and error are shown and could be used for better display. With regard to number of layers, we tried using several layers and found that using two layers produce the best results.

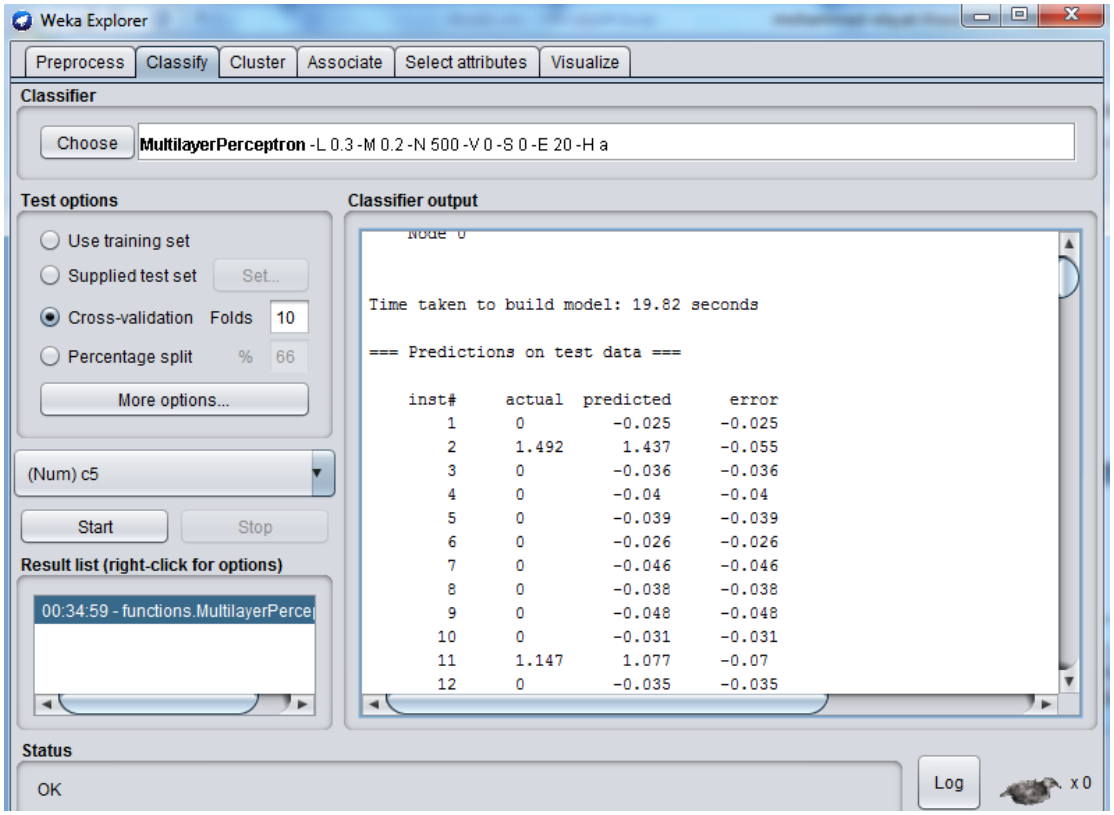


Figure 14 - Results with summary of statistics console.

6.5 Suggested Improvement Method of spatial analysis and data mining

Usually prediction depends on the points of data over the surrounding space, so the researcher also propose using the data of the neighbor points, which increases the number of variables considered for each observation point. Spatial data mining is suggested by taking into account the value of at least one variable value of the closest sites considering the closest neighbor areas; for example, it can be done considering the values of the rain. The logic behind this suggested implementation is the effect of the spatial space and variable on the handled observation. The suggested model is described by the following equation, which is demonstrated in Figure 15:

$$T_i = f\left(x_1, x_2, \ldots\ldots x_n, rain_{east}, rain_{west}, rain_{north}, rain_{south}\right)$$

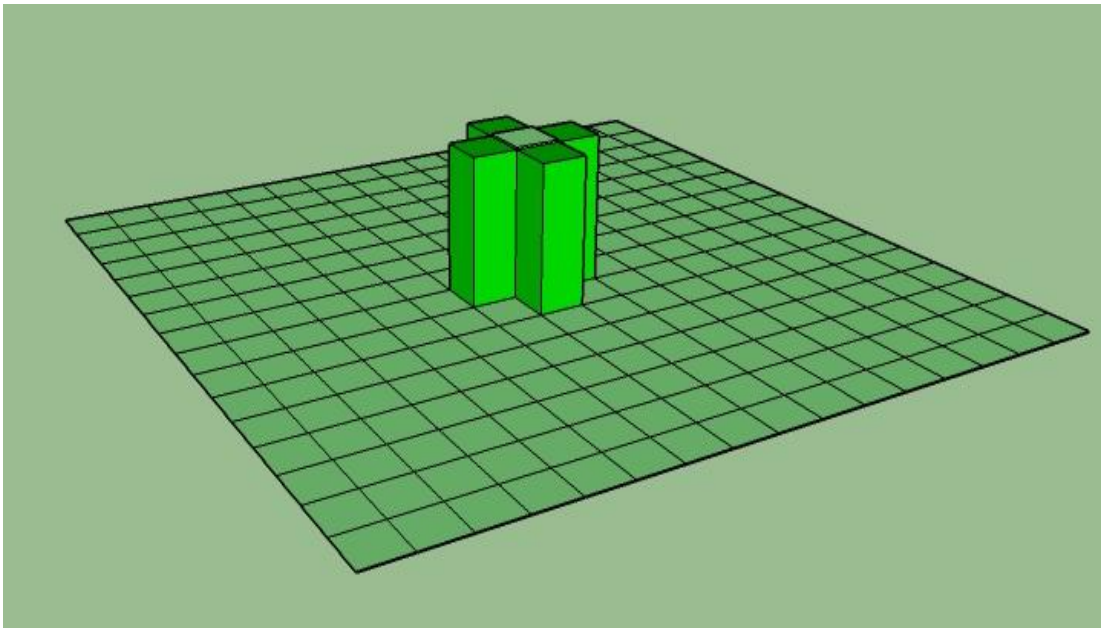


Figure 15- - Using the neighbors for each point over time

Thinking logically made this research improve the handled target value by also considering the target value of the neighbor sites and their effect on the handled point. However, the important question is: how can the calculated target value be considered while it's the target itself? And here it is suggested to start with the regular result values of the target considering the suggested variables. The process will be continued by iteration, while each iteration will consider the values from the previous one.

The suggested idea can be written by the following equation:

$$T_i^{k+1} = f\ (x_1, x_2, \dots\dots x_n, N_{west}^k, N_{east}^k, N_{north}^k, N_{south}^k)$$

where N is the neighbors values and k is the iteration index, and the suggested neighbors can be 4 or 8. Spatial analysis and modeling will deal with the area as a raster as mentioned before, Raster will be calculated and treated as a matrix dealing with map algebra. The suggested neighbors will be detected in the matrix according to the number of raw and column as shown in Figure 16.

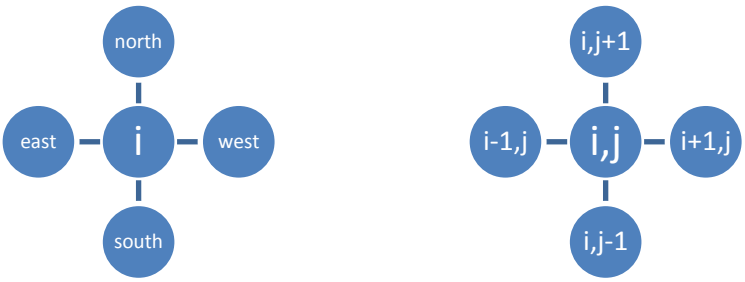


Figure 16 - The option of neighbors and their location as can be considered in the spatial

So the equation will be modified to the following form:

$$T_{i,j}^{k+1} = f\ (x_1, x_2, \dots\dots x_n, N_{i-1,j}^{k}, N_{i+1,j}^{k}, N_{i,j+1}^{k}, N_{i,j-1}^{k})$$

to arrange the data in a table , while each observation is presented in a record, and insert the values of the considered neighbors a vb.net code to read data from table and store it in a matrix (25, 34), then we calculate the neighbors for each point and store four neighbors at each point in the table using access database. Notice that eight neighbors can be considered by the same way.

The borders limitation and consideration:

Each point has four neighbors except the points at the edge of the raster, so we used two techniques to fix this issue: the first technique discards the points at the edge of the raster (minimize the raster by two columns and two rows). The other technique is to consider the neighbors in the borders a copy of the original point so that they won't affect the prediction result. The pseudo code below shows the researchers suggested algorithm.

Start

Initialize raster to1

Initialize i to1

Initialize j to1

While raster < 29

Read data from access database for the current raster

Store the data in matrix M(I,j)

determine point neighbors M(I,j+1), M(I,j-1), M(i+1,j), M(i-1,j)

For i=1 to 25

Forj=1 to 34

If M(I,j) point has the four neighbors then:

Determine the neighbors and store the data in

the matrix M(I,j)

else

Determine which neighbors are not found in the matrix

and consider its value as the same of the original point.

and store the data in the matrix M(I,j)

end for

end for

End while

Stop

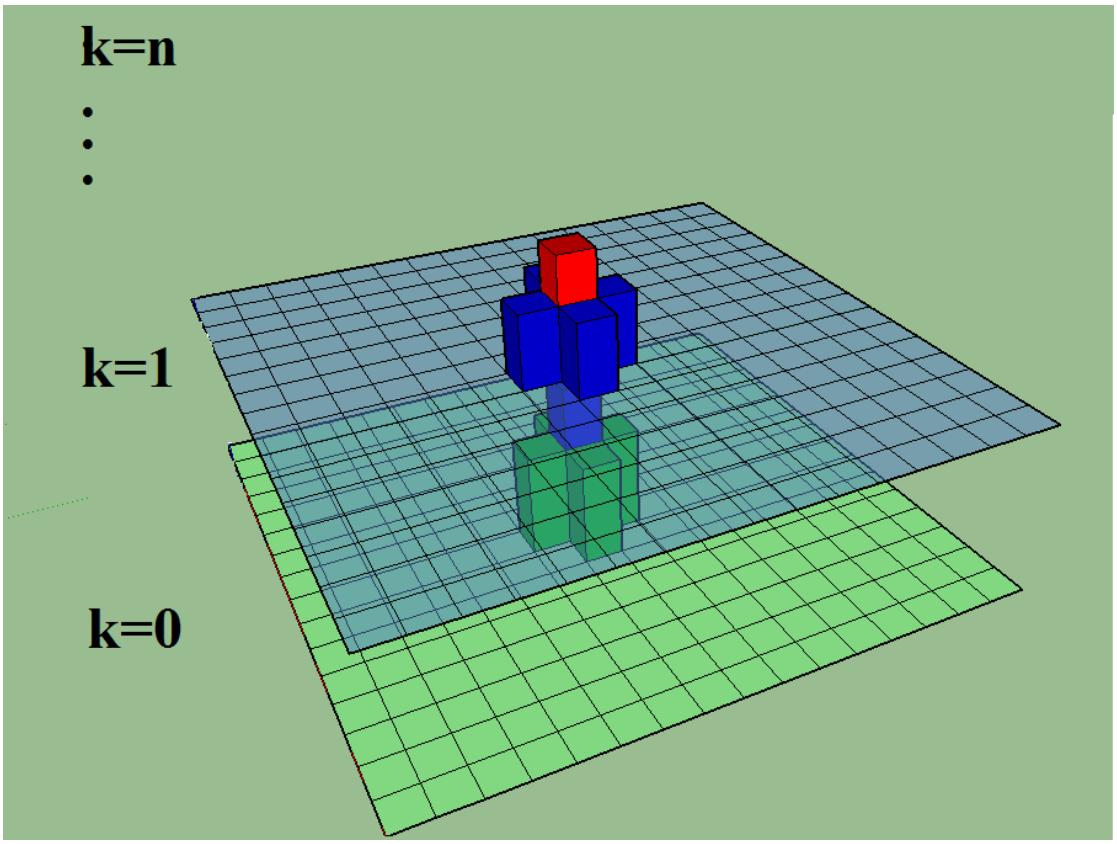
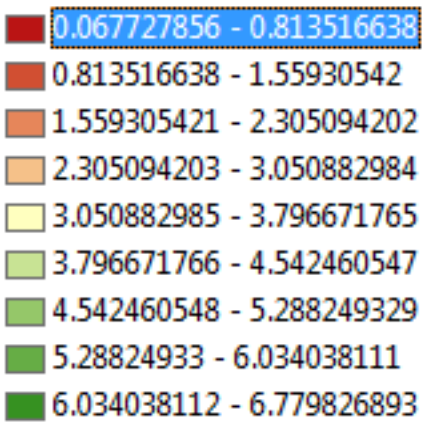


Figure 17 - using K iteration of the target values (affected percentage) to get more accurate results with the neighbors of the target value in iteration k+1.

The implementation of the suggested model of spatial data mining will be applied by neural network method. The maximum difference between each following iterations will be examined and will be treated as an indicator of the improvement of the results.

7. Results and Discussions

After using IDW interpolation for data, ArcGIS created a raster image for the layer of Jenin map, which represented the estimated data for Jenin area and is used to generate data for 850 points (25x34) based on the original 11 points. Each raster was converted into 850 and the estimated weather data are stored in a new table as a matrix. This number came as a result of choosing a point for each 1000 pixels as shown in Figures from no. 19 to 53.



IDW interpolation create a raster image for layer of Jenin map, each raster contain ninth classes as explained in figure 18 each class represents the variable effect in all locations that in our study area.

Figure 18 - shows IDW classes that represents weather variables effects in our study area

Figures from 19-25, the figures represents the IDW rain rasters in 2015 for January, February march april. Also figures from 22 to 24 represnts Affected Percentage Number of Affected items and Number of Affected persons in February/2015.

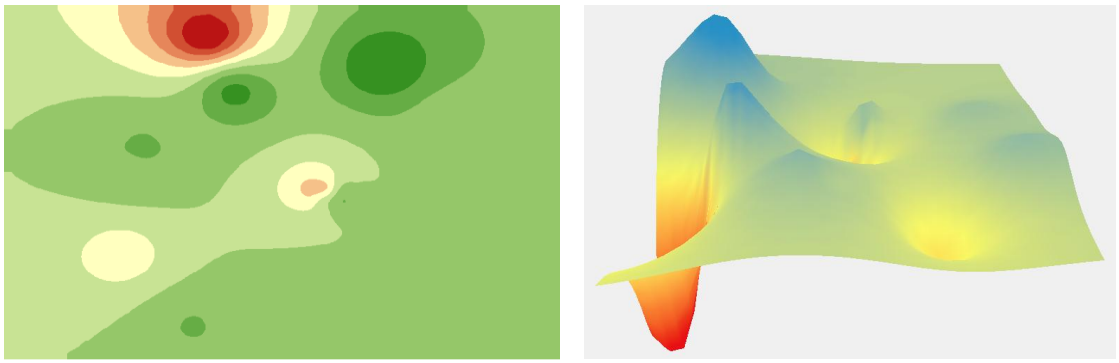


Figure 19 - IDW Rain raster in April / 2015

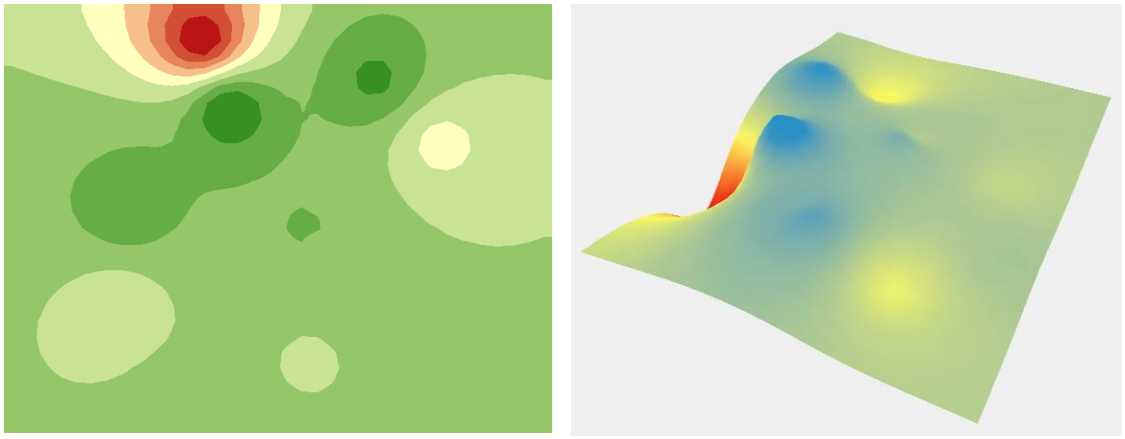


Figure 20- IDW Rain raster in March / 2015

Figure 21- IDW Rain raster in February / 2015

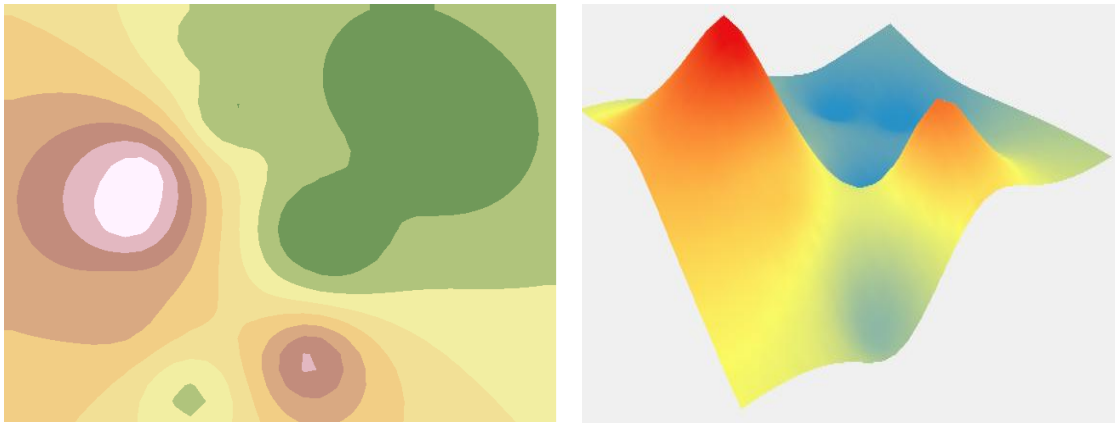


Figure 22- Number of Affected items in February / 2015

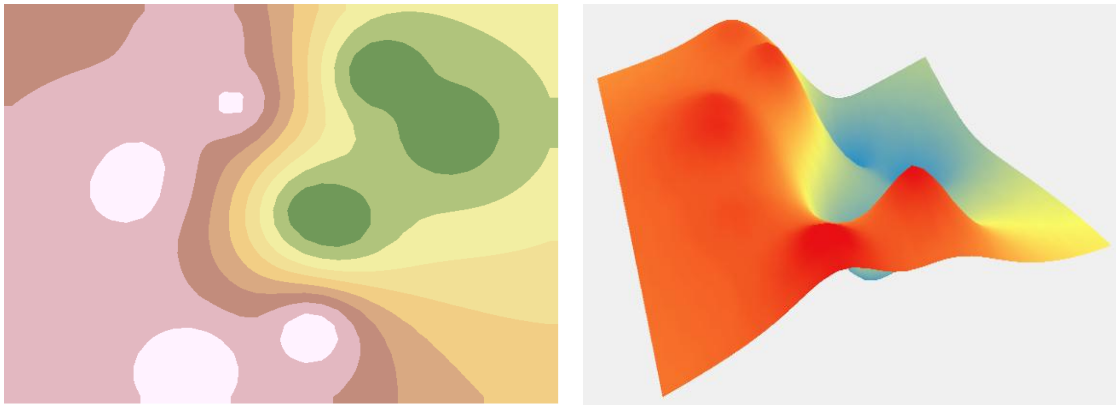


Figure 23- Affected Percentage in February / 2015

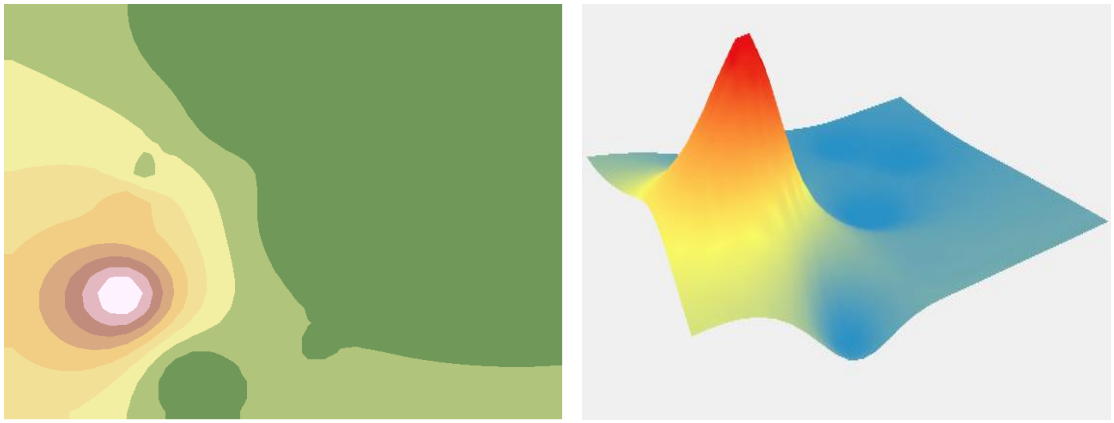


Figure 24- Number of Affected persons in February / 2015

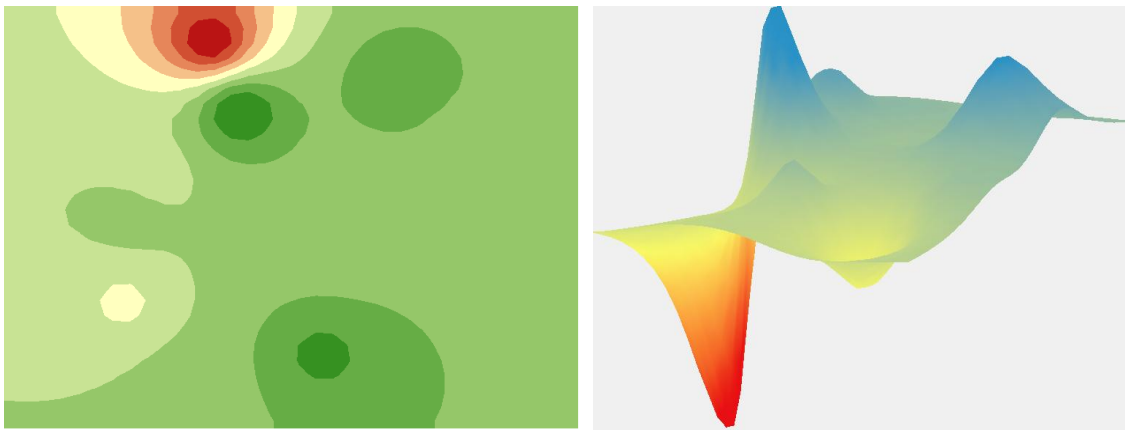


Figure 25-IDW rain raster in January / 2015

Figures from 26-35, the figures represents the IDW rain rasters in 2014 for January, February, march, may, october, November, December . Also figures from 31 to 33 represnts Affected Percentage Number of Affected items and Number of Affected persons in march /2014.

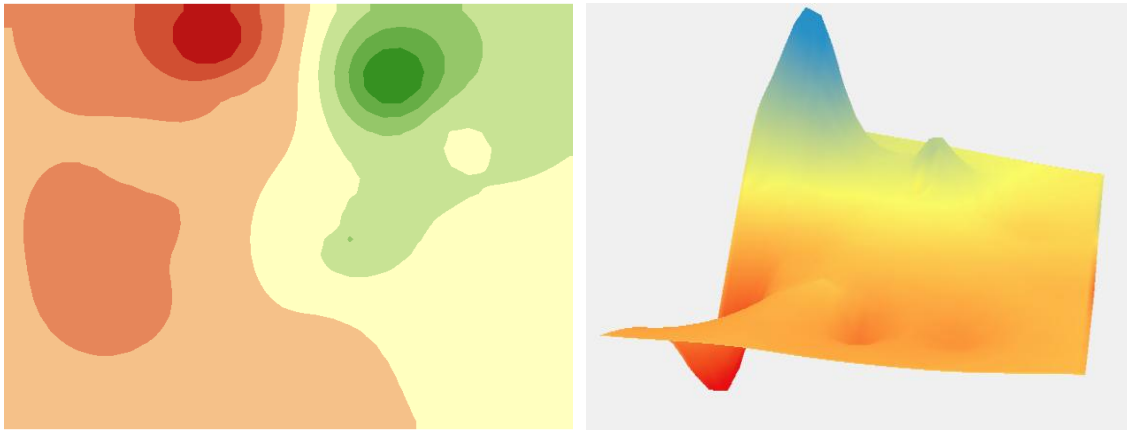


Figure 26- IDW rain raster in December / 2014

Figure 27- IDW rain raster in November / 2014

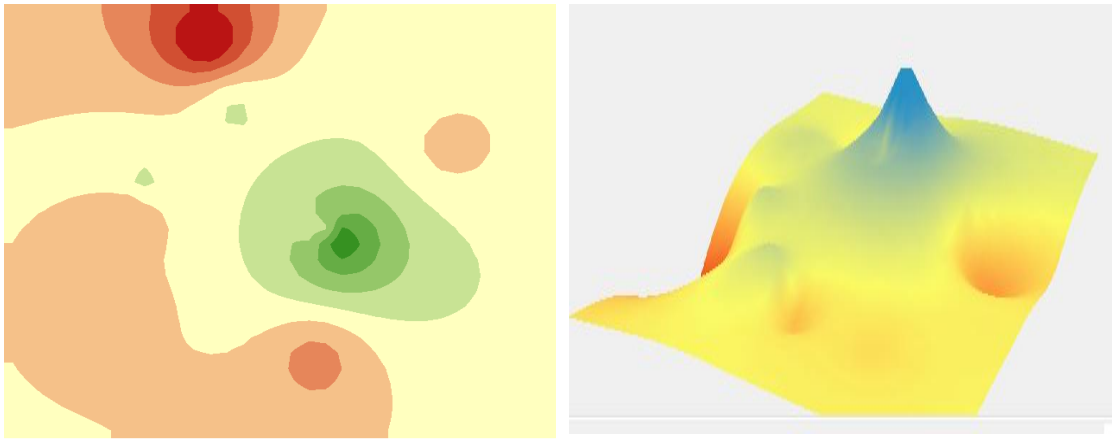


Figure 28- IDW rain raster in October / 2014

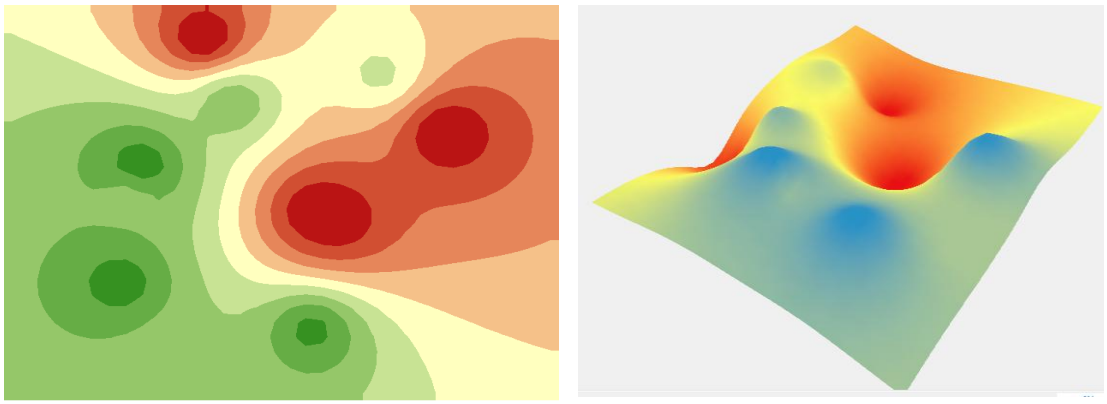


Figure 29 - IDW rain raster in May / 2014

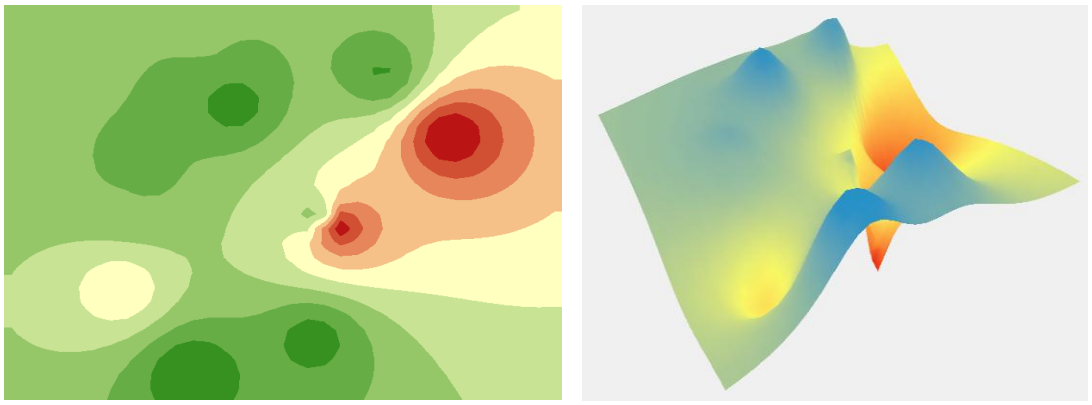


Figure 30 - IDW rain raster in march / 2014

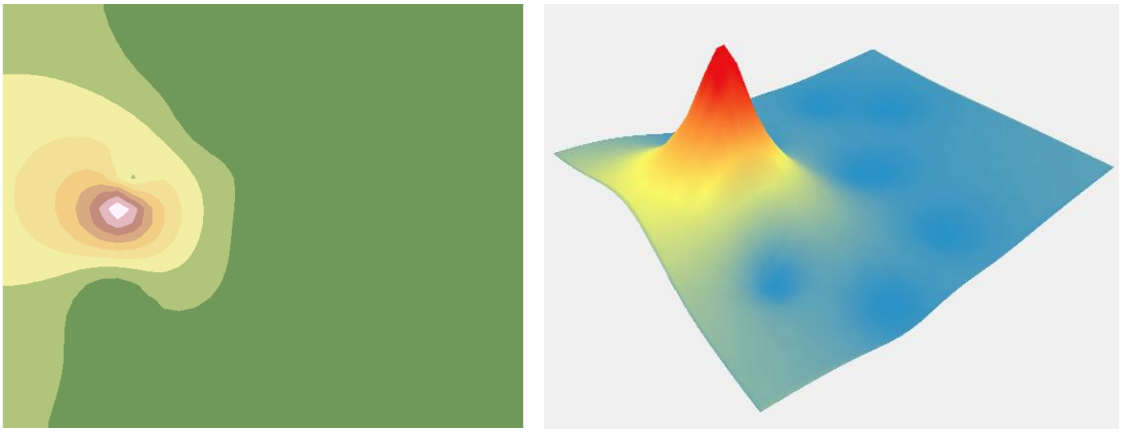


Figure 31 - IDW number of affected items raster in march / 2014

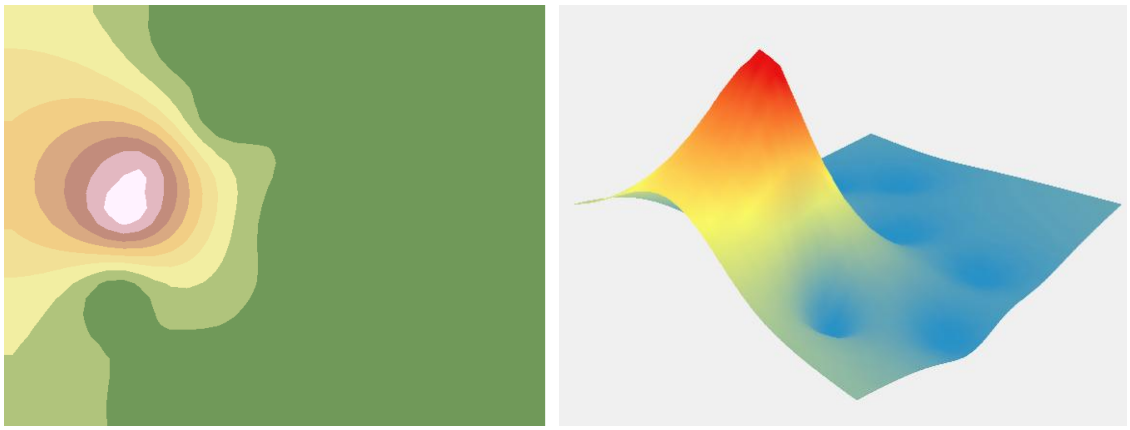


Figure 32 - IDW Affected percentage raster in March / 2014

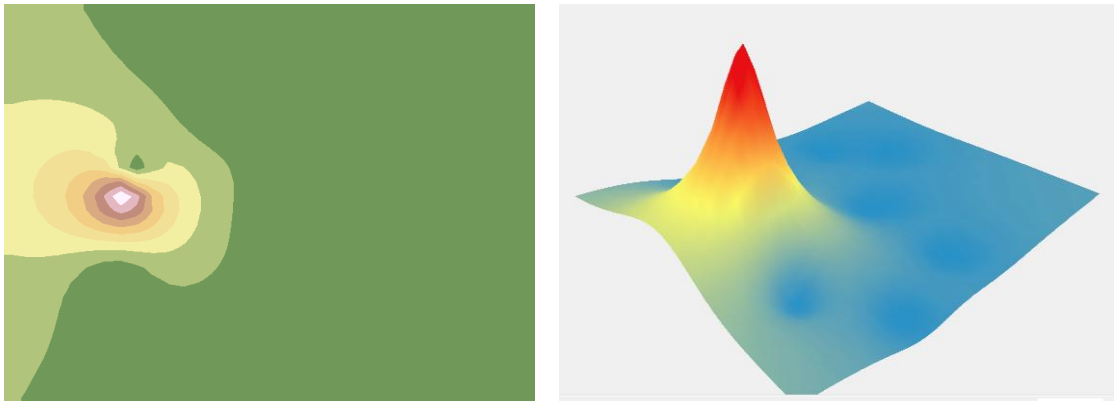


Figure 33 - IDW number of affected persons raster March / 2014

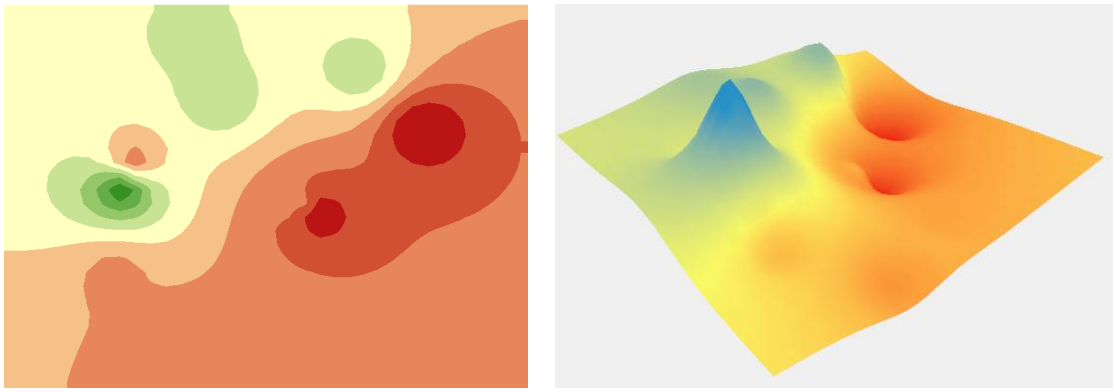


Figure 34 - IDW rain raster in February / 2014

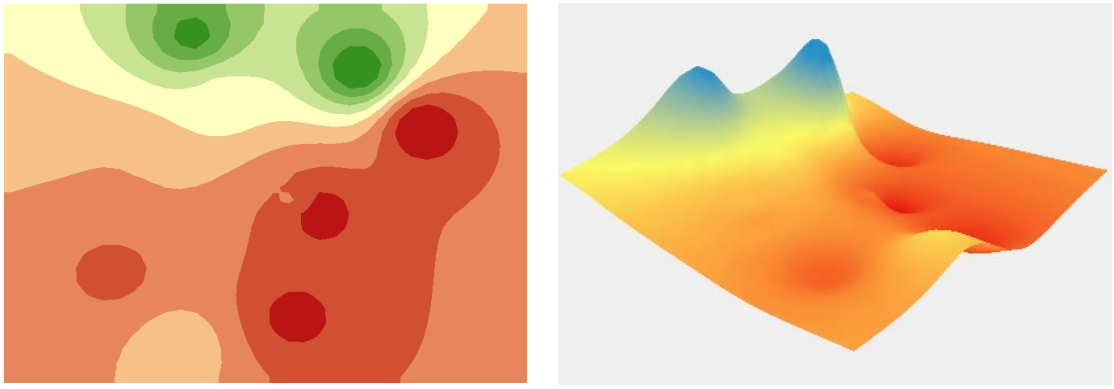


Figure 35 - IDW Rain raster in January / 2014

Figures from 36-44, the figures represents the IDW rain rasters in 2013 for January, February, march, april, october, November, December . Also figures from 42 to 44 represnts Affected Percentage Number of Affected items and Number of Affected persons in january /2013.

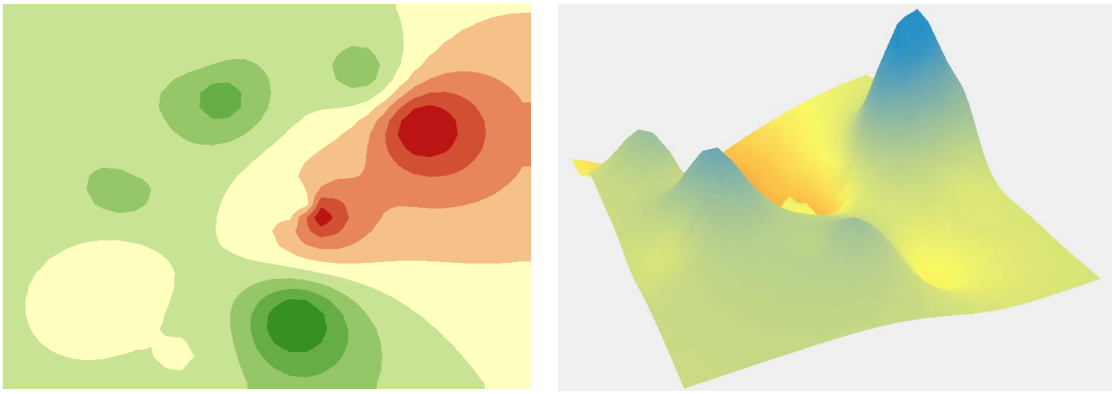


Figure 36 - IDW rain raster in December / 2013

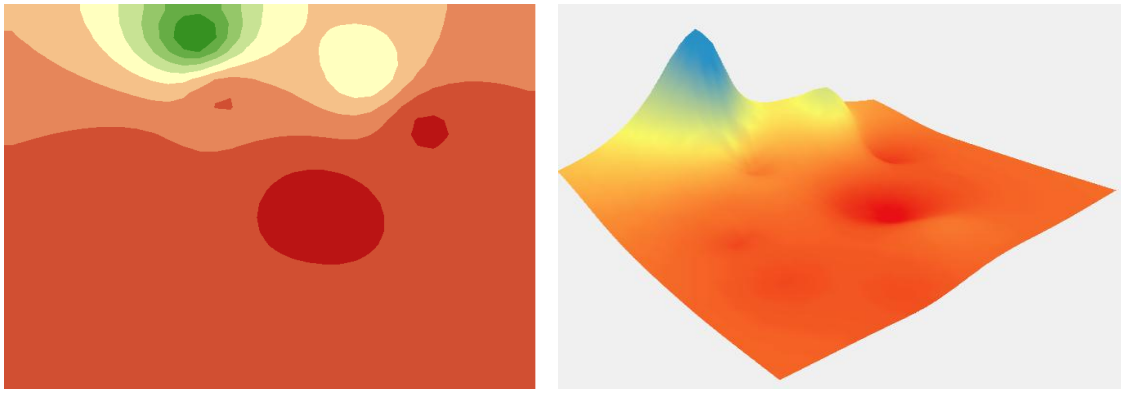


Figure 37 - IDW rain raster in November / 2013

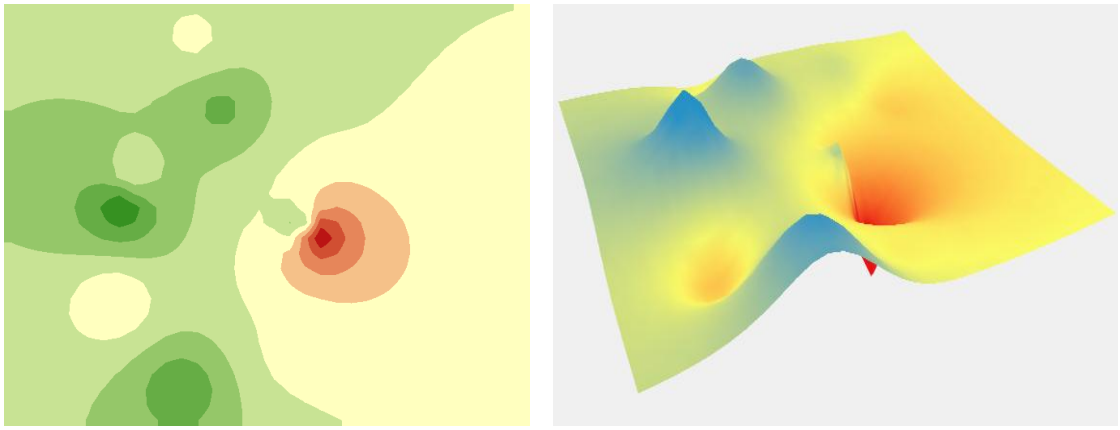


Figure 38 - IDW rain raster in April / 2013

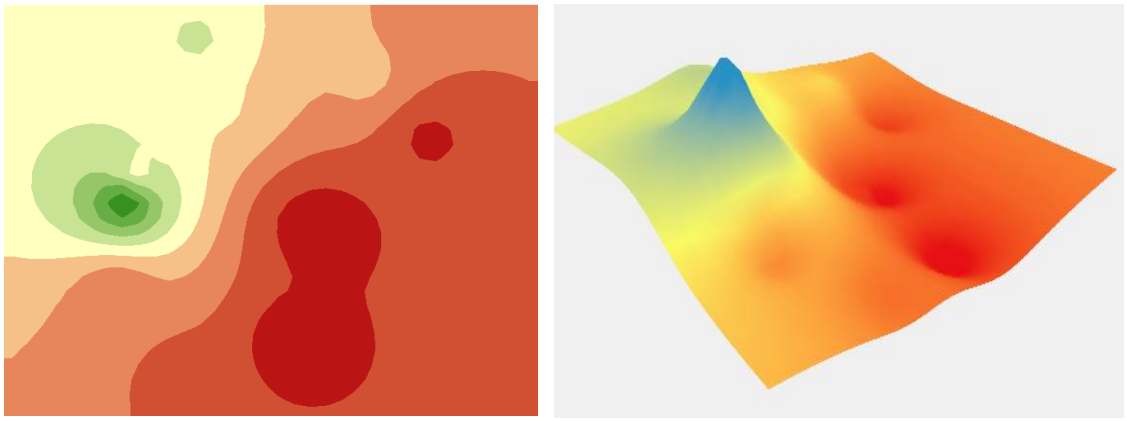


Figure 39 - IDW rain raster in March / 2013

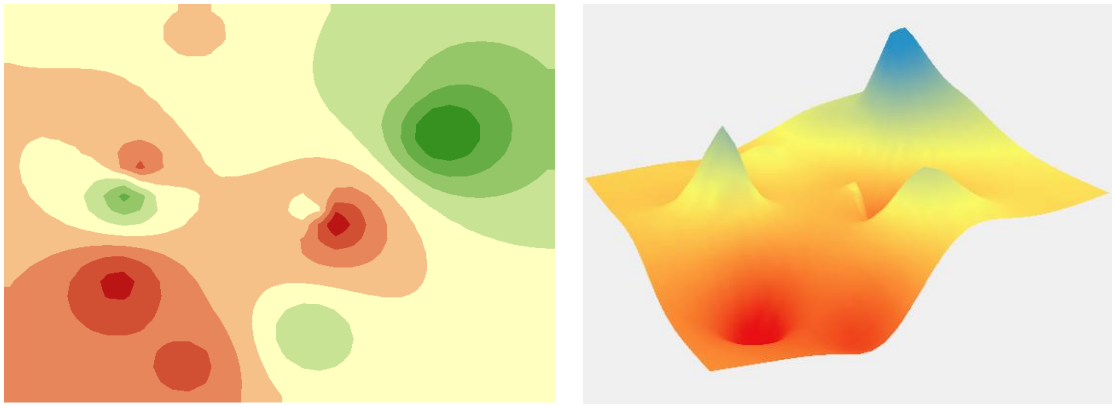


Figure 40 - IDW rain raster in February / 2013

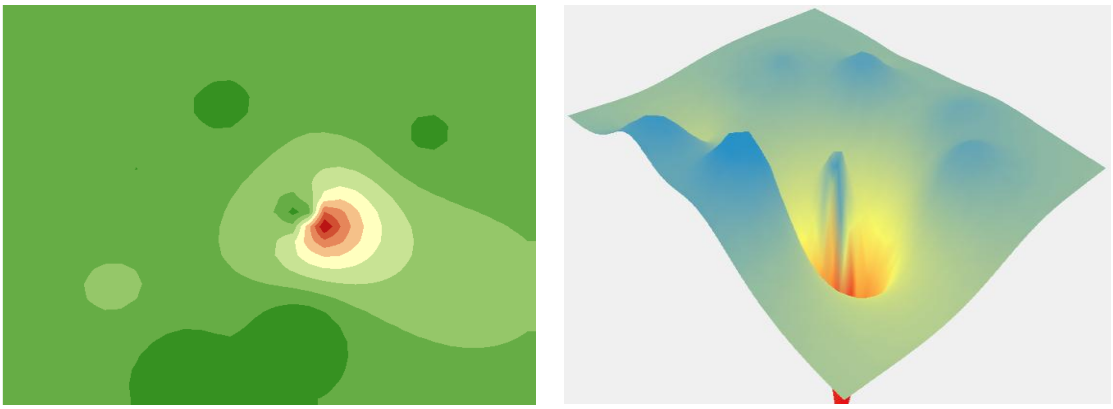


Figure 41 - IDW rain raster in January / 2013

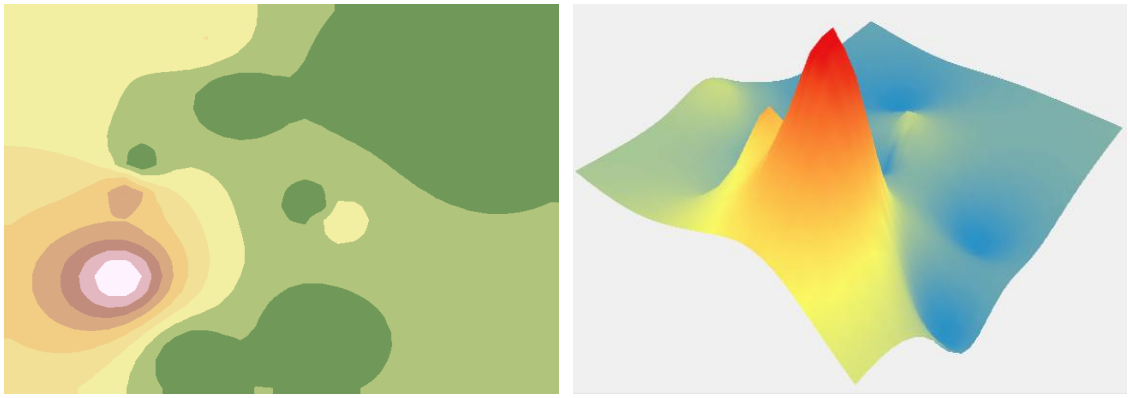


Figure 42 - IDW number of affected items in January / 2013

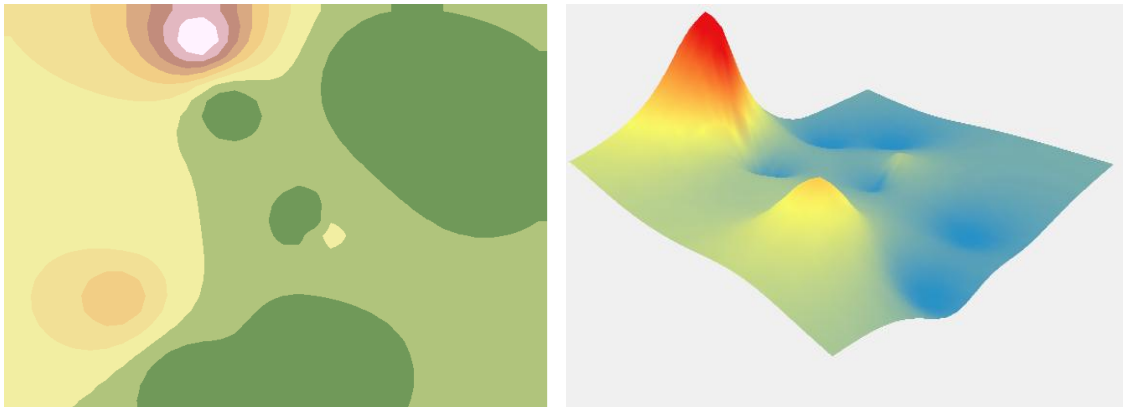


Figure 43 - IDW Affected percentage raster in January / 2013

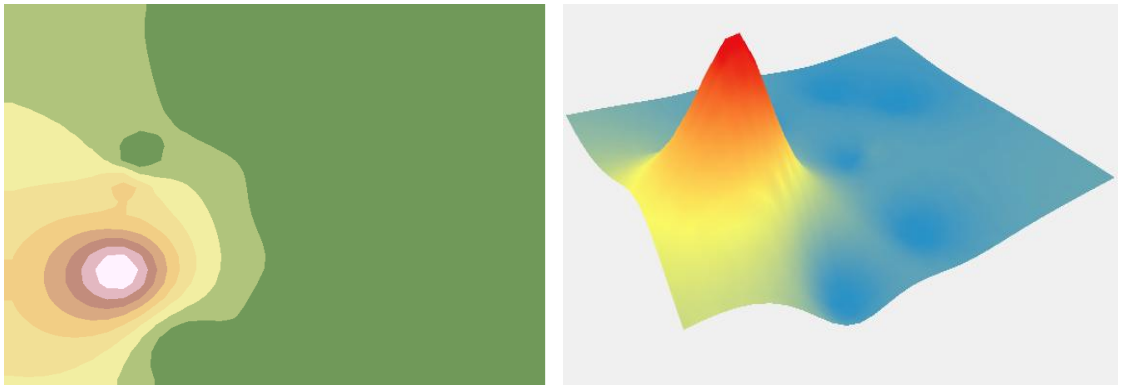


Figure 44 - IDW number of affected persons in January / 2013

Figures from 45-53, the figures represents the IDW rain rasters in 2012 for January, February, march, october, November, December . Also figures from 50 to 52 represnts Affected Percentage Number of Affected items and Number of Affected persons in February /2012.

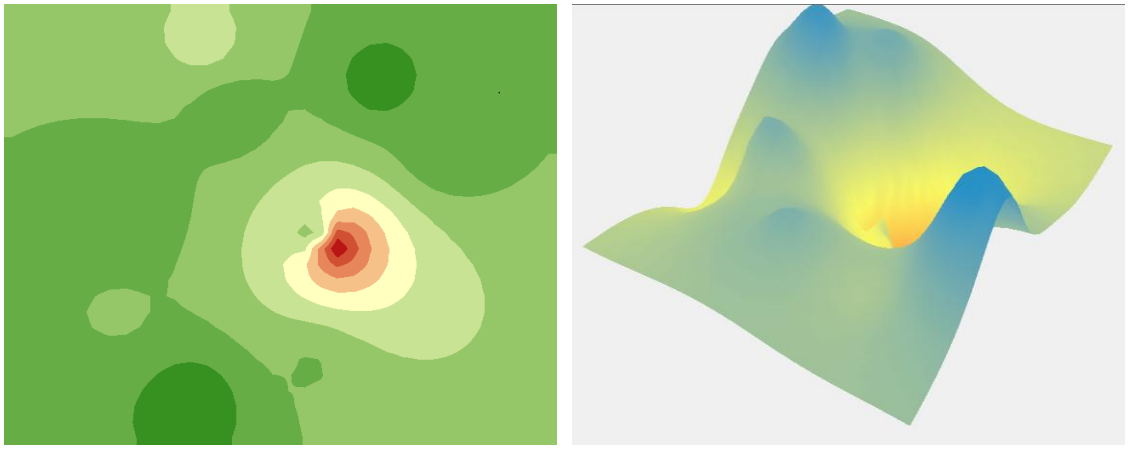


Figure 45- IDW Rain raster in December / 2012

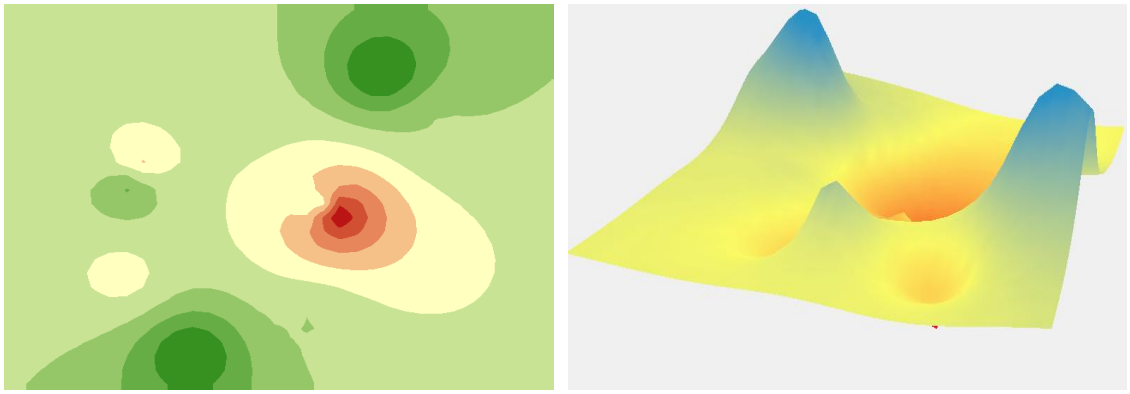


Figure 46- IDW rain raster in November / 2012

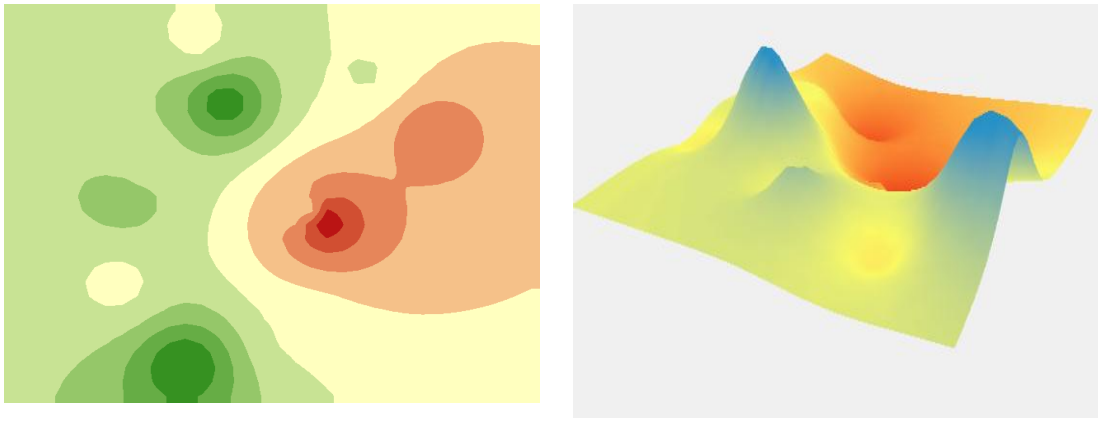


Figure 47- IDW rain raster in October / 2012

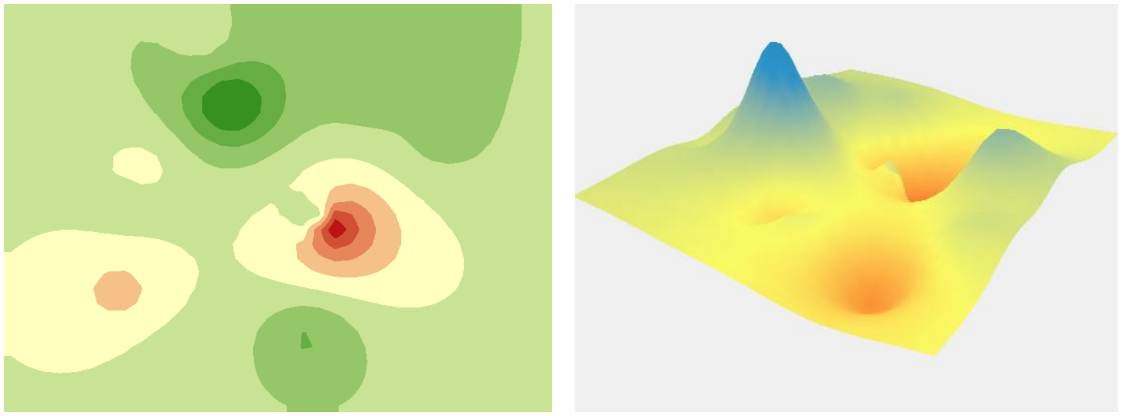


Figure 48- IDW rain raster in March / 2012

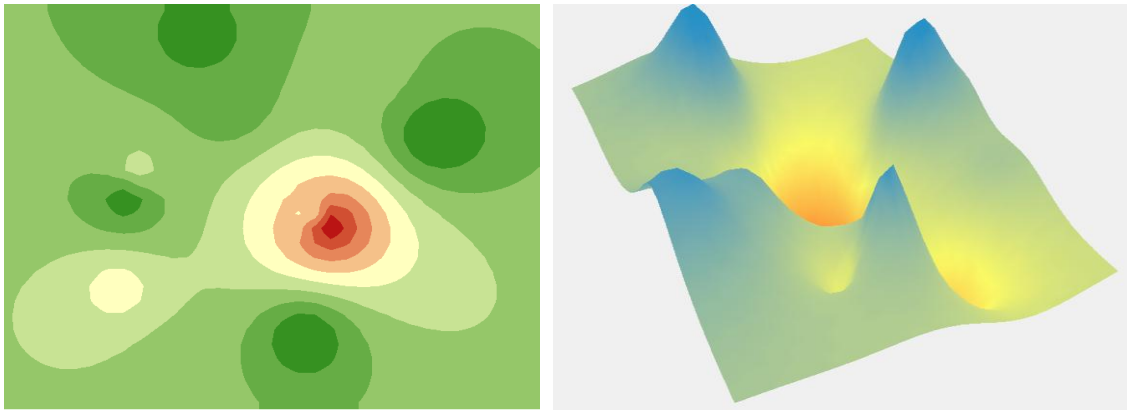


Figure 49- IDW rain raster in February / 2012

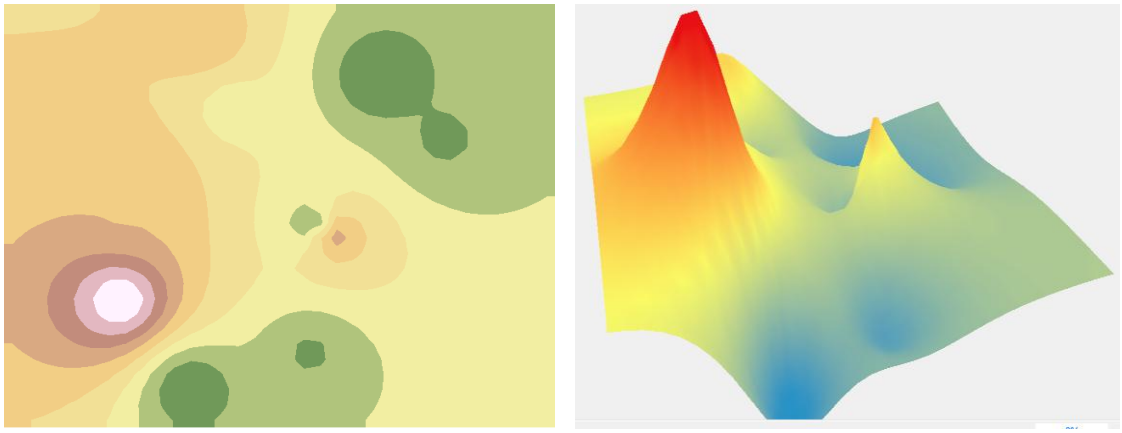


Figure 50- IDW number of affected items in February / 2012

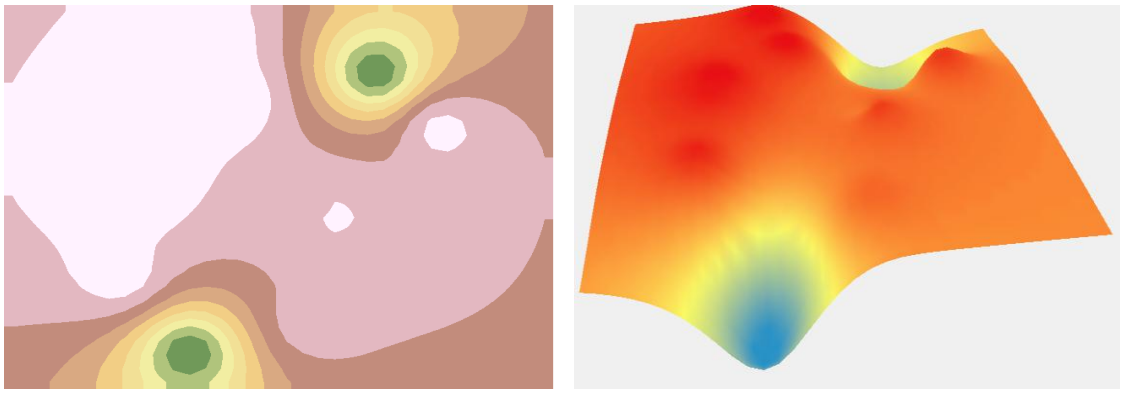


Figure 51- IDW Affected percentage raster in February / 2012

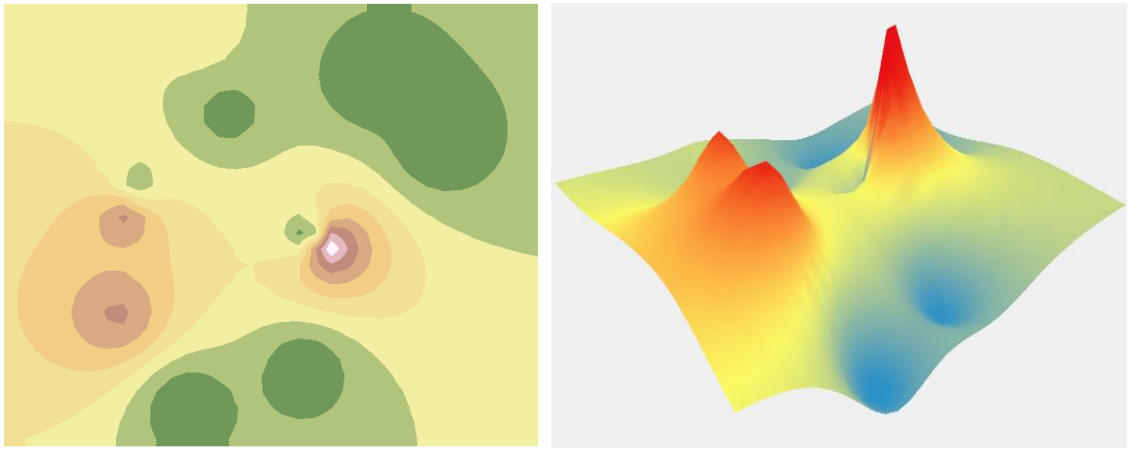


Figure 52- IDW number of affected persons raster in February / 2012

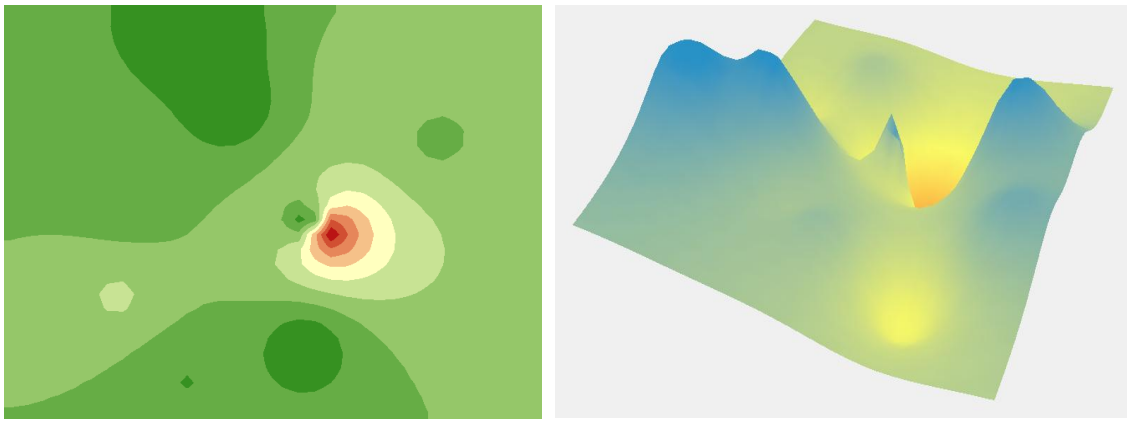


Figure 53- IDW rain raster in January / 2012

3D analysis tools and theories were used to understand, analyze and simulate the elevation and the slope for the analyzed area and surface to be considered and added to the handled variables. The elevation as shown in figure no. 54 can be calculated by interpolation theories. Jenin area, as a study case, was converted into 3D raster surface by using contours of 5m (contour is a line that connects the points with the same elevation) as a topographic map by the tool Topo to Raster. Advanced 3D analysis of the surface elevation can calculate each observation point of the slope of the topographic surface by angels see Figure no. 55.

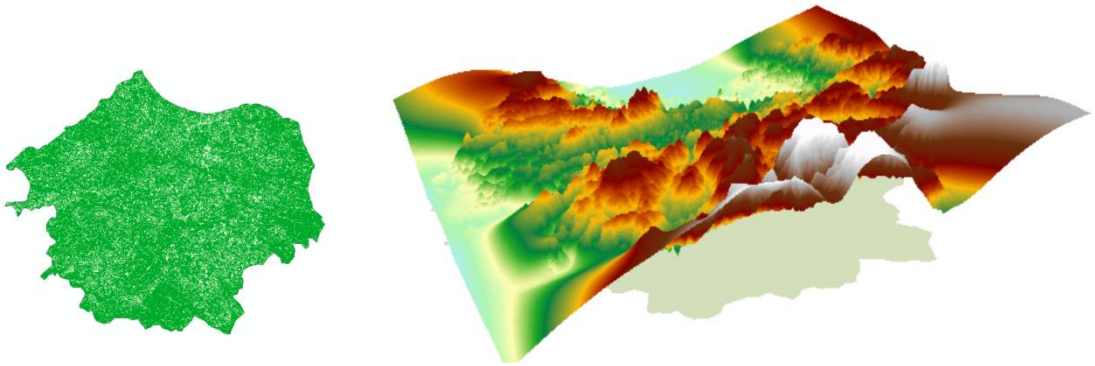


Figure 54- Converting topographic contours into 3D surface and raster of Jenin elevations (the image with z factor for elevations)

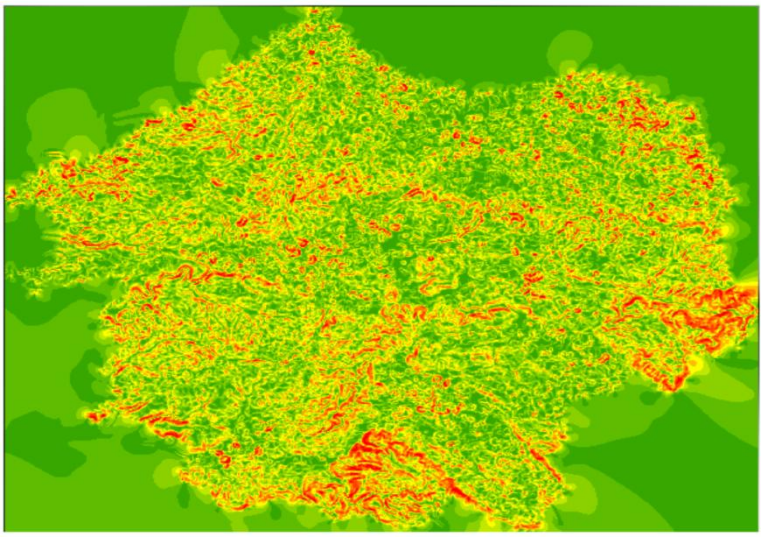


Figure 55- Raster output file that shows the calculation of the slope of the topographic surface in each point

After applying Ordinary Least Square (OLS) to data without elevation and without slope, the result appears in Figure no. 56.

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	105.007491	4.176874	25.140209	0.000000*	4.545727	23.100263	0.000000*	--------
WIND	-0.440104	0.056569	-7.779931	0.000000*	0.059973	-7.338315	0.000000*	1.049990
EVAP	0.008248	0.078387	0.105223	0.916182	0.028664	0.287748	0.773550	1.392002
SUNSHINE	2.934471	0.171896	17.071213	0.000000*	0.142152	20.643236	0.000000*	7.467478
RH	-8.516049	0.361345	-23.567645	0.000000*	0.389047	-21.889504	0.000000*	7.656098
TEMPAV	-7.208456	0.123329	-58.449128	0.000000*	0.139042	-51.843655	0.000000*	3.783703
RAIN	5.490151	0.078410	70.018760	0.000000*	0.092205	59.542599	0.000000*	2.330674

Figure 56- A summary of OLS results – Model variables without elevation and slope

After entering elevation attribute and applying Ordinary Least Square (OLS), the result appears in Figure no. 57.

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	104.880092	4.186157	25.054028	0.000000*	4.555289	23.023807	0.000000*	--------
WIND	-0.440136	0.056570	-7.780375	0.000000*	0.059972	-7.339021	0.000000*	1.049992
EVAP	0.008188	0.078388	0.104449	0.916797	0.028665	0.285625	0.775176	1.392006
SUNSHINE	2.934614	0.171899	17.071743	0.000000*	0.142147	20.644877	0.000000*	7.467502
RH	-8.516496	0.361352	-23.568420	0.000000*	0.389040	-21.891079	0.000000*	7.656154
TEMPAV	-7.208365	0.123331	-58.447373	0.000000*	0.139053	-51.839046	0.000000*	3.783713
RAIN	5.490593	0.078417	70.017985	0.000000*	0.092250	59.518687	0.000000*	2.331027
ELEVA	0.031026	0.067609	0.458898	0.646327	0.066599	0.465858	0.641336	1.000151

Figure 57- summary of OLS results – Model variables

After we had added a slope attribute and applied Ordinary Least Square to new data, the researcher got new results shown in Figure no. 58. And then results in the three stages were compared and shown in Table 2.

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	107.050709	4.228019	25.319355	0.000000*	4.611669	23.213006	0.000000*	--------
WIND	-0.475005	0.057373	-8.279177	0.000000*	0.059607	-7.968918	0.000000*	1.080555
EVAP	-0.001000	0.078410	-0.012754	0.989819	0.028386	-0.035231	0.971884	1.393471
SUNSHINE	2.907356	0.172022	16.901039	0.000000*	0.142957	20.337329	0.000000*	7.481888
RH	-8.653365	0.363244	-23.822481	0.000000*	0.391833	-22.084310	0.000000*	7.740308
TEMPAV	-7.230118	0.123448	-58.568339	0.000000*	0.139558	-51.807441	0.000000*	3.792732
RAIN	5.475753	0.078505	69.750189	0.000000*	0.092088	59.462302	0.000000*	2.337423
SLOP	-0.250676	0.069369	-3.613666	0.000317*	0.064025	-3.915282	0.000100*	1.065074
ELEVA	0.051274	0.067825	0.755978	0.449660	0.066993	0.765361	0.444053	1.007024

Figure 58 - summary of OLS results – Model variables

There are several variables used to measure the OLS result. The most important measures are mentioned below:

First, Robust probabilities, which are used to decide if a variable is helping our model or not. Second ,the coefficient for each explanatory variable reflects both the strength and type of relationship between explanatory variable and dependent variable. Coefficient negative value means that the relationship is negative; on the other hand, when the sign is positive, the relationship is positive. The unity of the coefficients are shown in the same units as their associated explanatory variables. Third, the T test is used to check that an explanatory variable is statistically significant or not. [41]

The fourth is the VIF measures explanatory variables redundancy. As a rule, if the explanatory variables associated with VIF values are larger than about 7.5, they should be removed from the regression model. In summary, these histograms clarify the relationship between the dependent variable and each explanatory variable.

We could also present the 3D chart in Figure 63 that shows Estimated, Residual and error for Ordinary Least Square after adding elevation and slope attributes.

The comparison between the results of the ordinary least square with or without elevation and slope calculations is given in the table2 below. It's clear that these variables didn't affect the results that were given without considering theses spatial variables in this case. These variables won't be considered in the next implementations.

The considered variables	Adjusted Residuals
Without elevation and without slope	0.365347
With slope and with elevation	0.365637
With elevation and Without slope	0.365327

Table 2- comparison between the results of the Ordinary Least Square with or without elevation and slope

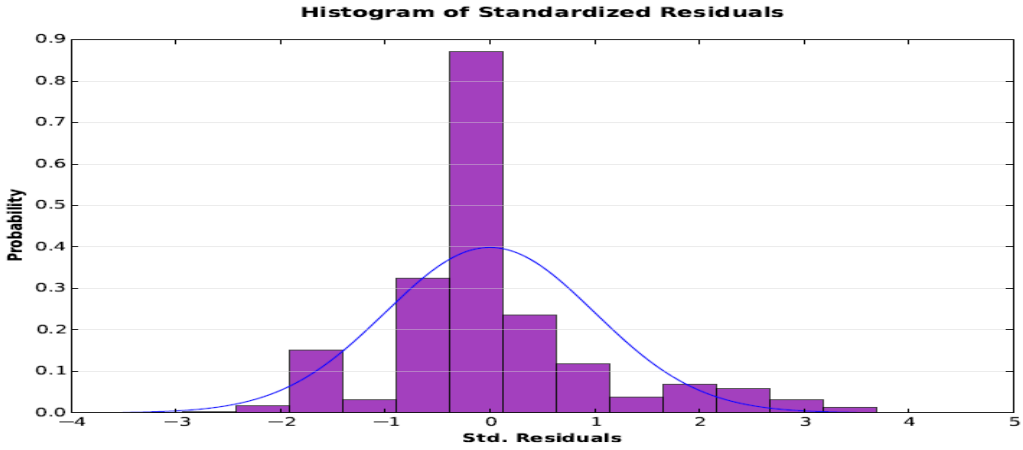


Figure 59- residuals with elevations and with slope

In the figure of Histogram of Standardized residuals, there are blue lines which show the shape of the histogram if the residuals will take normal distribution. And the histogram bars shows the distribution.

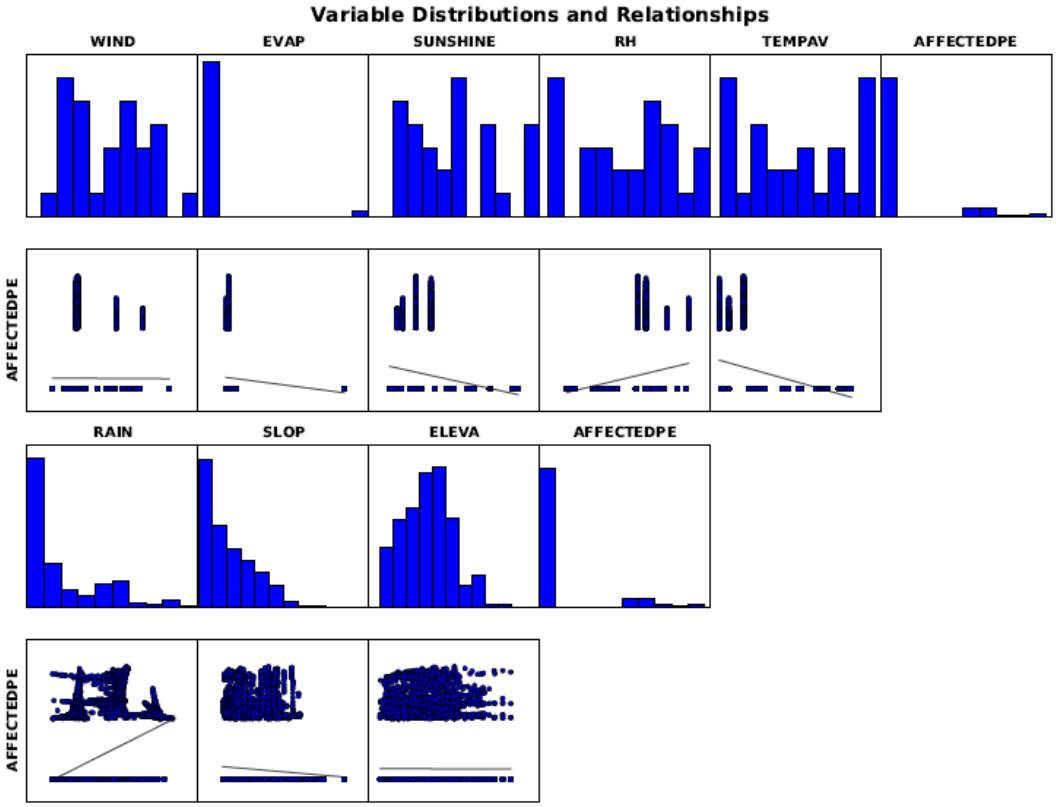


Figure 60 - variable distributions and relationships

The above graphs are Histograms for each explanatory variable and the dependent variable. Each variable shows how it is distributed.

Each graph shows the relationship between an explanatory variable and the dependent variable. Diagonal shapes represent the strong relationships which is either positive or negative.

Figure 61 - graph of error percentage between standard residuals and predicted values in 2013 and 2014

Applying a training set and multilayer perceptron network method for the original data before interpolation got good enough results as shown in Figure 62 below.

Correlation coefficient	0.995
Mean absolute error	1.3837
Root mean squared error	2.6203
Relative absolute error	8.2148 %
Total Number of Instances	352

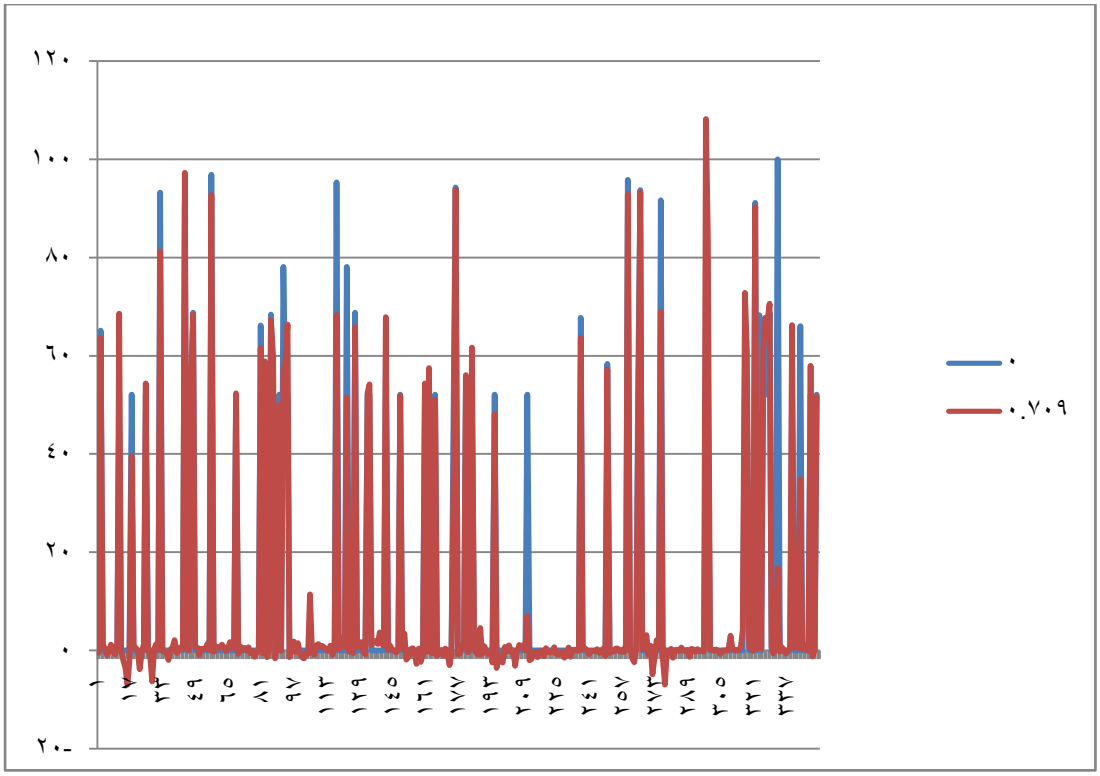


Figure 62 - Applying K-fold Cross validation and multilayer perceptron network for data before interpolation

The researcher applied IDW interpolation for the data and then applied the cross validation and multilayer perceptron neural network to get more accurate results in predictions, and after applying this method, the researcher noted that prediction results are better than before interpolation as we see in Figure 63 below.

Correlation coefficient	0.9959
Mean absolute error	0.5695
Root mean squared error	1.0606
Relative absolute error	3.703 %
Total Number of Instances	24650

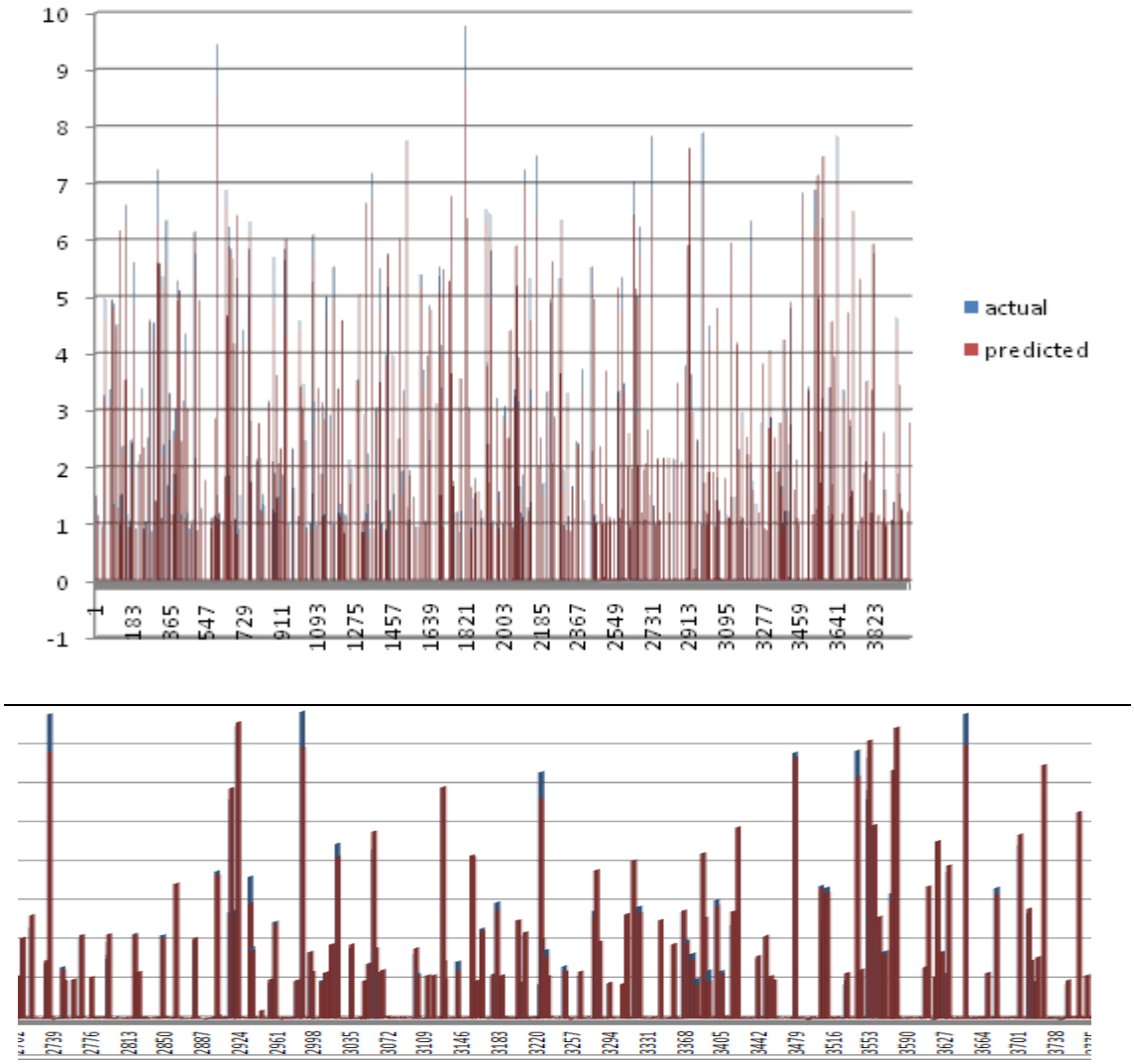


Figure 63 - Applying K-fold Cross validation and multilayer perceptron network for data after interpolation.

After applying multilayer perceptron network and using cross validation for the data after interpolation with four neighbor points around each point, the prediction result was much better than taking each point with data changing over time.

Also, we can reach better results using K =0,1,2,3. We decided that we don't need bigger values of K because the mean absolute error was satisfactory as shown in Table3.

K=0 prediction results neural network without K iteration

$T_{k\,(}i,j) = F($ rain, wind, temp, ……., $t_{k-1}(i-1,j)$, $t_{k-1}(i+1,j)$, $t_{k-1}(i,j+1)$, $t_{k-1}(i,j-1))$

The table 3 below is presenting the results of the prediction for all different ways that we used.

	Correlation coefficient	Mean absolute error	Root mean squared error	Relative absolute error	Total Number of Instances
Original points before interpolation	0.995	1.3837	2.6203	8.2148 %	352
After interpolation	0.9969	0.6919	1.8439	4.4989 %	24650
K0 with neighbors	0.9989	0.3773	1.1287	2.4534 %	24650
K1 with neighbors	0.9996	0.2467	0.6324	1.6246 %	24650
K2 with neighbors	1	0.0726	0.1768	0.4802 %	24650
K3 with neighbors	1	0.0718	0.1664	0.3714%	24650

Table 3 - result of prediction for all different techniques

K0 maximum error = 14.377
K1 maximum error = 6.817
K2 maximum error = 1.651
K3 maximum error = 1.251

As we see in the above table , the accuracy of the results has increased after applying data interpolation and the researcher's proposed approach (taking four neighbors of each one observation) has got better results in comparison with the other known techniques.

8. Conclusion and Future Work

This thesis analyzed the climate change impact on agricultural productivity in Jenin due to the importance of the climatic conditions and their impact on the agricultural production; the necessary data was collected and digitized, such as wind speed, temperature, humidity, rainfall, air pressure etc. The data collected in hard copy tables from the meteorological station and the Agriculture Directorate in Jenin Governorate was the main goal to predict the agricultural risks.

The conclusions drawn from this study indicate that after collecting the data and preparing it for analysis and modeling, several operations were applied in data preprocessing , such as normalization and data interpolation.

The importance of these phases were clear in the results improvements and the spatial covering of all the points of the researcher's study area in predicting agricultural damage. Whereas other researchers applied some preparations for input data, also we applied more preparations on our input data, which is very important to achieve better results. In this research spatial location and topology relationships were consider in order to ensure success and effective analysis to these spatial environment phenomena. The suggested implementation in this thesis built a connection between adjacent points to evaluate their effect and consider the dual influence. The suggested advanced consideration caused to reduce the error rate in our work, moreover each one of the researchers use one algorithm or one model for prediction and didn’t make comparison with algorithms, so we applied several theories and developed new ideas in prediction, also we make comparisons between theories we used and choose what theory brought us better results.

It was noticed that the error rate was reduced after using data Interpolation. However, taking the adjacent points of inputs, such as considering the rainfall and estimating the

agricultural damage as a target increased the accuracy of the results. Maybe the Ordinary least square was not the perfect theory to be applied, but, at least, it shows an equation with the coefficient that highlights the effect of each variable. On the other hand, the complexity of neural network theory and modeling, which brought us better results, makes it hard to be presented as a model.

We are aware of the sensitivity of these theories to the data, its type and the preparing process. The limitation of data and time justified the application of interpolation with such distance and without any information about the data sources and accuracy, but the aim of this research is to highlight the effect of GIS and spatial implementation on the data mining domain.

The results of each implementation were examined with the given values; residuals of testing data and cross validation were tested. Satisfying results were achieved and presented. A clear improvement was achieved by the suggested spatial model in two levels when neighbor's variables were considered and furthermore when the target value themselves of the neighbor areas were considered using iterations. The results of the iterations were tested by calculating the maximum difference between the current and previous iteration and consequently an impressive improvement was proved.

Our future work in this thesis deals with the data that need to be more accurate and real. That is, they should be taken in a higher resolution to ensure the correctness of the interpolation implementations. More data can be considered, such as the land use, type of soil, fertilizer and yield, maybe the underground water and other variables that can be very dominant and we didn’t have in this research.

A 3D analysis was done to calculate and consider the elevation and the slope of each area, but an aspect analysis can be applied(see figure no. 66) and can consider the direction of each surface area.

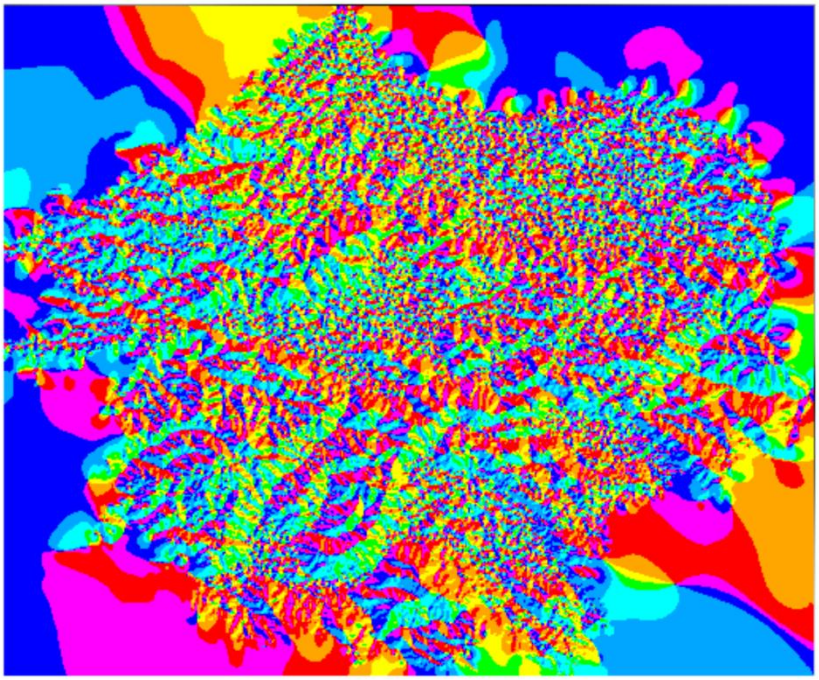


Figure 64 -The results of aspect analysis on Jenin 3D surface

The suggested iteration spatial model can be faster convergent and easily implemented if it is applied and programmed in Mathlab or R programming language, where the neural network function can be run immediately and then values from the current iteration for the earlier observation will be considered. Notice the red values in the suggested equation:

$$T_{i,j}^{k+1} = f\ (x_1, x_2, \dots\dots\dots x_n, N_{i-1,j}^{k+1}, N_{i+1,j}^{k}, N_{i,j+1}^{k}, N_{i,j-1}^{k+1})$$

Time space data mining was examined, but it failed because of the lack of data for long time range, whereas spatial time data mining can be applied to a time cube and

space or even 4D while considering the Euclidian space and other parametric space and several serials of time.

9. References

1. www.iasri.res.in/ebook/win_school_aa/notes/spatial_data_mining.pdf Retrieved on 1.03.2017

2. Eldrandaly, K.: Expert systems, GIS, and spatial decision making: current practices and newtrends. In: Expert Systems: Research Trends, pp. 207–22 (2007)

3. Ladner, R., Petry, F.E., Cobb, M.A.: Fuzzy set approaches to spatial data mining ofassociation rules. Trans. GIS 7(1), 123–138 (2003)

4. Prediction of outcome of construction dispute claims using multilayer perceptron neural network model N.B. Chaphalkar a , K.C. Iyer b , Smita K. Patil (2015).

5. Baboo S., and Shereef K., "Applicability of Data Mining Techniques for Climate Prediction – A Survey Approach," International Journal of Computer Science and Information Security, Vol. 8, No. 1, April 2010.

6. https://opensource.com/alternatives/arcgis Retrieved on 2.05.2017

7. http://www.isprs.org Retrieved on 20.07.2017

8. http://pro.arcgis.com/en/pro-app/tool-reference/space-time-pattern-mining /creat-space-time-cube.htm Retrieved on 15.04.2017
9. http://www.icaci.org Retrieved on 10.03.2017

10. Spatial Data Mining: Progress and Challenges Survey paper Krzysztof KoperskiJunasAdhikaryJiawei Han fkoperski, adhikary, hang@cs.sfu.ca School of Computing Science Simon Fraser University Burnaby, B.C., Canada V5A 1S6

11. Elmasri, R. and Navathe, S. 2010, Fundamentals of Database Systems, Addison-Wesley Publishing Company.

12. Chaphalkar, N.B., K.C. Iyer, and Smita K. Patil. "Prediction of outcome of construction dispute claims using multilayer perceptron neural network model" , International Journal of Project Management, 2015.

13. Mucherino, A., Papajorgji, P., & Pardalos, P. (2009), “Data mining in agriculture” (Vol. 34), Springer.

14. (2012)A hybrid neural networks-fuzzy logic-genetic algorithm for grade estimation Author links open overlay panelPejmanTahmasebiArdeshirHezarkhani

15. A Survey on Data Mining Techniques in Agriculture M.C.S.Geetha Assistant Professor, Dept. of Computer Applications, Kumaraguru College of Technology, Coimbatore, India.(2015)

16. https://medium.com/@xenonstack/overview-of-artificial-neural-networks-and-its-applications-2525c1addff7 Retrieved on 15.08.2017

17. Renewables 2015 Global Status Report, 2015

18. http://www. research.ijcaonline.org Retrieved on 20.09.2017

19. Ministry of New and Renewable energy, Government of India,“Annual Report 2015-16”, http://mnre.gov.in, 2016.

20. https://www.researchgate.net/post/Best_normalization_techniques Retrieved on 1.11.2017

21. Normalization: A Preprocessing Stage S.Gopal Krishna Patro1 , Kishore Kumar sahu2 Research Scholar, Department of CSE & IT, VSSUT, Burla, Odisha, India

22. https://www.quora.com/What-is-the-meaning-of-min-max-normalization Retrieved on 10.12.2017

23. Spatial and Temporal Variations of Dissolved Oxygen in ChaAm Municipality Wastewater Treatment Ponds Using GIS Kriging Interpolation Shwesin Koko,1 Kim N. Irvine,2 Ranjna Jindal1 and Romanee Thongdara1 1 Mahidol University, Thailand; 2 Nanyang Technology University, Singapore

24. Using a hybrid methodology of dasyametric mapping and data interpolation techniques to undertake population data (dis)aggregation in South Africa*Mawande Ngidi, Gerbrand Mans, David McKelly, Zukisa Sogoni*

25. Interpolation Techniques and Associated Software for Environmental DataArjunAkkala,a Vijay Devabhaktuni,a and Ashok Kumarba EECS Department, The University of Toledo, Toledo, OH 43606b Department of Civil Engineering, The University of Toledo, Toledo, OH 43606;

26. http://statweb.stanford.edu/~tibs/sta306bfiles/cvwrong.pdf Retrieved on 10.10.2017

27. Cross-validation of the Student Perceptions of Team-Based Learning Scale in the United StatesDonald H. Lein, Jr,1 John D. Lowman,1,* Christopher A. Eidson,2 and Hon K. Yuen2

28. http://www.deeplearning.net/tutorial/mlp.html Retrieved on 05.09.2017

29. https://www.analyticsvidhya.com/learning-paths-data-science-business-analytics-business-intelligence-big-data/weka-gui-learn-machine-learning/ Retrieved on 1.11.2017

30. Elmasri, R. and Navathe, S. 2010, Fundamentals of Database Systems, Addison-Wesley Publishing Company.

31. Laila Mohamed ElFangary, Maryam Hazman, AlaaEldinAbdallahYassin" Mining the Impact of Climate Change on Animal Production".

32. Vale, M. M., Moura, D. J., Naas, I. de A., Oliveira, S. R. de M., and Rodrigues, L. H. A. 2008. Data Mining to Estimate Broiler Mortality When Exposed to Heat Wave, Sci. Agric. (Piracicaba, Braz.), Vol. 65, No. 3, 223-229.

33. Sarah N. Kohail, Alaa M. El-Halees" Implementation of Data Mining Techniques for Meteorological Data Analysis "

34. Swati Hira, P.S. Desh pande. "Data Analysis Using Multidimensional Modeling Statistical Analysis and Data Mining on Agriculture Parameter", Procedia Computer Science, Vol.54, pp: 431-439, 2015

35. Harln D. Shannon, Raymond P. Motha. "Managing Weather and Climate Risk to Agriculture North America, Central America and the Caribbean", Vol. 10, pp: 50-56, December 2015

36. disasters SaptarsiGoswami a , Sanjay Chakraborty a, *, SanhitaGhosh a , AmlanChakrabarti b , BasabiChakraborty c "A review on application of data mining techniques to combat natural disasters" Faculty of Software and Information Science, Iwate Prefectural University, Japan

37. FolorunshoOlaiya" Application of Data Mining Techniques in Weather Prediction and Climate Change Studies" Department of Computer & Information Systems, Achievers University, Owo, Nigeria.

38. Mallari, C.Alyosha, Ezra. "Climate change Vulnerability Assessment in the Agriculture Sector: Typhon Santi Experience", Procedia- Social and Behavioral Sciences, Vol. 260, pp: 440-451, January 2016

39. https://www.google.ps/search?q=jenin+map&source=lnms&tbm=isch&sa=X&ved=0ahUKEwib6MPe5sPZAhWjh6YKHXt5DagQ_AUICigB&biw=1242&bih=602#imgrc=2hoDcyA6Qpq6yM: Retrieved on 12.03.2017

40. https://www.support.esri.com/en/technical-article/000008595 Retrieved on 15.10.2017

41. http://resources.esri.com/help/9.3/arcgisengine/java/gp_toolref/spatial_statistics_tools/interpreting_ols_results.htm Retrieved on 12.11.2017

42. Irrigated and Dry Farming in Jenin Governorate Plains (Comparative Study) Prepared by Nahed Mahmoud Rafeq Zakarneh (2012)

43. http://www.xenonstack.com Retrieved on 1.10.2017

44. https://opensource.com/alternatives/arcgis Retrieved on 15.09.2017

45. Kanna Bhaskar and S.N. Singh,"AWNN - Assisted Wind Power Forecasting using FeedForward Neural Network". IEEE Transactions on Sustainable Energy, Volume 3, pp. 306- 315, 2012.

البيانات والتنبؤ بنسبة المحاصيل المتضررة قبل وبعد استيفاء البيانات. حيث تم تطبيق هذه الخوارزميات باستخدام عدة برامج من ضمنها برنامج نظم المعلومات الجغرافية ArcGIS وبرنامج WEKA.

في المرحلة الاخير من هذه الاطروحة تم اقتراح طريقة جديدة للتنبؤ المكاني بحالة الطقس باستخدام خريطة الجبر والتي تعامل منطقة الخرائط كما أنها مصفوفة من البيانات حيث تم كتابة كود برمجي وذلك لتحويل البيانات، وفرزها وتطبيق عمليات التنقيب عن البيانات المكانية الجديدة. تم تطبيق النهج المكاني من خلال النظر في النقاط الأربعة المحيطة لكل نقطة لضمان الأخذ بعين الاعتبار تأثير هذه المناطق المحيطة على نسبة الضرر. تم تطبيق مختلف الأساليب على بيانات التدريب واختبارها من خلال عملية cross validation. وتم حساب النتائج ومقارنتها مع النتائج التي حصلنا عليها قبل المقترح والتي قدمت نتائج مرضية حيث تم مقارنة جميع العمليات التي تم تنفيذها حيث حصل النهج المقترح على أفضل النتائج للتنبؤ بحالة الطقس والاضرار المتعلقة بالزراعة، هذا المقترح والنتائج التي حصلنا عليها تشجع المزيد من البحوث في هذا المجال ويقترح العمل به في المستقبل بما يتضمن من تحليل متقدم ونمذجة.

10. ملخص الرسالة باللغة العربية

إن الظروف الجوية بما في ذلك هطول الأمطار، ودرجة الحرارة، وسرعة الرياح، والرطوبة تؤثر على نمو النباتات والزراعة بطرق مختلفة حيث يمكن أن تجعلها أكثر عرضة للاصابة بالأمراض وتعطي فرصة لنمو الحشرات. كما أن الظروف المناخية القاسية يمكن أن تشكل الخطر الاكبر على نمو النباتات فمن الممكن ان تقتل النباتات او خلال ساعات وبالتالي الحاق الضرر بالمزارعين مما قد يؤدي إلى خسارة زراعية ضخمة، لذلك عملية التنبؤ بالطقس هو واحد من أهم القضايا العلمية والتكنولوجية ويعتبر تحدي من تحديات مشاكل القرن.

في هذه الأطروحة، تم تطبيق أساليب استخراج البيانات المكانية على بيانات الطقس التي تم جمعها من أجل التنبؤ بالأحوال الجوية وتنبيه المزارعين في وقت مبكر لاخذ جميع الاحتياطات التي من غرضها تقليل الخسائر الزراعية. البيانات التي تم جمعها تشمل العديد من عوامل الطقس من كميات الأمطار ودرجة الحرارة وسرعة الرياح والرطوبة والارتفاع عن سطح البحر ونسبة المحاصيل المتضررة من عدة مصادر من هيئة الأرصاد الجوية الفلسطينية ووزارة الزراعة.

أجريت عدد من العمليات على البيانات التي تم جمعها بعدة مراحل، حيث تم ترقيمها، تحضيرها وتنظيفها من البيانات الغير ضرورية وعمل تقريب للقيم ضمن مدى واحد لضمان التحليل والحصول على نتائج صحيحة، ثم العمل على تحويل البيانات إلى تنسيقات تناسب نظم المعلومات الجغرافية ، حيث تم استخدام برنامج ArcGIS، تم استخدام برنامج ArcGIS لحفظ البيانات وعرضها بطريقة يسهل فهمها، وأيضا تم استخدامه لربط البيانات مع مواقعها الجغرافية على مدار الزمن ضمن فترات زمنية محددة، وبالتالي استخدام عدد من الخوارزميات لتحليل البيانات. كما تم استخدام عملية (Interpolation) على البيانات لاستيفائها أي تم تمديد البيانات بحيث تغطي جميع النقاط التي تبعد حوالي 1000 متر عن بعضها البعض وبالتالي الحصول على بيانات تشمل منطقة الدراسة المطلوبة. وقد تم تنفيذ العديد من الخوارزميات، مثل ordinary least square and multi perceptron neural network وذلك لاستخراج

```
Imports System.Data.OleDb
Public Class Form1
   Private Sub Button1_Click(sender As Object, e As EventArgs) Handles Button1.Click
      Dim x, y As Integer
      x = 34
      y = 25

      Dim matrix(x, y) As Double

      For i = 1 To 29
         Label1.Text = i
         Dim j As Integer

         Dim conn As OleDbConnection
         Dim sql, str As String
 str = "Provider=Microsoft.ACE.OLEDB.12.0;Data Source=D:\newdata.accdb"
         conn = New OleDbConnection(str)
         sql = " Select rain from newdata where  id = " & i & ""
         Dim cmd As OleDbCommand
         Dim r As OleDbDataReader
         conn.Open()
         cmd = New OleDbCommand(sql, conn)
         r = cmd.ExecuteReader
         If r.HasRows Then
```

```
    For x = 1 To 34
        For y = 1 To 25
            r.Read()
            matrix(x, y) = r(0)
        Next
    Next

    conn.Close()
    OleDbConnection.ReleaseObjectPool()
Else
    conn.Close()
    MsgBox("nodata")
End If

Dim rain1, rain2, rain3, rain4 As Double

For x = 1 To 34
    For y = 1 To 25
        If x = 1 And y = 1 Then
            rain1 = matrix(x + 1, y)
            rain2 = matrix(x, y)
            rain3 = matrix(x, y)
            rain4 = matrix(x, y + 1)
        ElseIf x = 34 And y = 1 Then
            rain1 = matrix(x, y)
            rain2 = matrix(x, y)
            rain3 = matrix(x - 1, y)
```

```
    rain4 = matrix(x, y + 1)
ElseIf x = 1 And y = 25 Then
    rain1 = matrix(x + 1, y)
    rain2 = matrix(x, y - 1)
    rain3 = matrix(x, y)
    rain4 = matrix(x, y)
ElseIf x = 34 And y = 25 Then
    rain1 = matrix(x, y)
    rain2 = matrix(x, y - 1)
    rain3 = matrix(x - 1, y)
    rain4 = matrix(x, y)
ElseIf y = 1 Then
    rain1 = matrix(x + 1, y)
    rain2 = matrix(x, y)
    rain3 = matrix(x - 1, y)
    rain4 = matrix(x, y + 1)
ElseIf x = 1 Then
    rain1 = matrix(x + 1, y)
    rain2 = matrix(x, y - 1)
    rain3 = matrix(x, y)
    rain4 = matrix(x, y + 1)
ElseIf y = 25 Then
    rain1 = matrix(x + 1, y)
    rain2 = matrix(x, y - 1)
    rain3 = matrix(x - 1, y)
    rain4 = matrix(x, y)
ElseIf x = 34 Then
```

```
                rain1 = matrix(x, y)
                rain2 = matrix(x, y - 1)
                rain3 = matrix(x - 1, y)
                rain4 = matrix(x, y + 1)

            Else
                rain1 = matrix(x + 1, y)
                rain2 = matrix(x, y - 1)
                rain3 = matrix(x - 1, y)
                rain4 = matrix(x, y + 1)
            End If

            Dim conn1 As OleDbConnection
            Dim sql1, str1 As String
str1 = "Provider=Microsoft.ACE.OLEDB.12.0;Data Source=D:\newdata.accdb"
            conn1 = New OleDbConnection(str1)
            sql1 = "update [newdata] set [rain1] = " & rain1 & " , [rain2] = " & rain2 & " ,
[rain3] = " & rain3 & " , [rain4] = " & rain4 & " where xcoo = " & x & " and ycoo = " & y &
" and id = " & i & " "
            Dim cmd1 As OleDbCommand
            cmd1 = New OleDbCommand(sql1, conn1)
            conn1.Open()
            cmd1.ExecuteNonQuery()
            conn1.Close()
            OleDbConnection.ReleaseObjectPool()
        Next
    Next
```

```
        Next
    End Sub
```

Vb.net code for collecting four neighbors of every point and store them in the table.

```
/**
 *
 * @author mohammed.eleyat
 */
import weka.core.Instances;
import weka.core.converters.ArffSaver;
import weka.core.converters.CSVLoader;
import java.io.File;
public class CSV2Arff {
  /**
   * takes 2 arguments:
   * - CSV input file
   * - ARFF output file
   */
  public static void main(String[] args) throws Exception {
  /*  if (args.length != 2) {
      System.out.println("\nUsage: CSV2Arff <input.csv><output.arff>\n");
      System.exit(1);
    }*/
    // load CSV
    CSVLoader loader = new CSVLoader();
```

```
    loader.setSource(new File("d:\\IInormalizedData1.csv"));
    Instances data = loader.getDataSet();
    // save ARFF
    ArffSaver saver = new ArffSaver();
    saver.setInstances(data);
    saver.setFile(new File("d:\\IInormalizedData1.arff"));
    saver.setDestination(new File("d:\\IInormalizedData1.arff"));
    saver.writeBatch();
  }
}
```

Java code to convert data in excel file to arff file