**Software Requirements Specification**

For

**Linear Regression Using Gradient Descent**

**9th November 2021**

Prepared by

|  |  |  |
| --- | --- | --- |
| **Specialization** | **SAP ID** | **Name** |
| AIML | 500075346 | Dhruv |
| AIML | 500076009 | Dipanshu |
| AIML | 500076448 | Ghanisht |
| AIML | 500076007 | Jasleen |



Department of Informatics

School Of Computer Science

UNIVERSITY OF PETROLEUM & ENERGY STUDIES,

DEHRADUN- 248007. Uttarakhand

Table of Contents

|  |  |  |
| --- | --- | --- |
| **Topic** | | **Page No** |
| Table of Content | |  |
| Revision History | |  |
| 1 | Introduction |  |
|  | 1.1 Purpose of the Project |  |
|  | 1.2 Target Beneficiary |  |
|  | 1.3 Project Scope |  |
|  | 1.4 References |  |
| 2 | Project Description |  |
|  | 2.1 Reference Algorithm |  |
|  | 2.2 Data/ Data structure |  |
|  | 2.3 SWOT Analysis |  |
|  | 2.4 Project Features |  |
|  | 2.5 User Classes and Characteristics |  |
|  | 2.6 Design and Implementation Constraints |  |
|  | 2.7 Design diagrams |  |
|  | 2.8 Assumption and Dependencies |  |
| 3 | System Requirements |  |
|  | 3.1 User Interface |  |
|  | 3.2 Software Interface |  |
|  | 3.3 Database Interface |  |
|  | 3.4 Protocols |  |
| 4 | Non-functional Requirements |  |
|  | 4.1 Performance requirements |  |
|  | 4.2 Security requirements |  |
|  | 4.3 Software Quality Attributes |  |
| 5 | Other Requirements |  |
| Appendix A: Glossary | |  |
| Appendix B: Analysis Model | |  |
| Appendix C: Issues List | |  |

**Revision History**

|  |  |  |  |
| --- | --- | --- | --- |
| **Date** | **Change** | **Reason for Changes** | **Mentor Signature** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**1: INTRODUCTION:**

Machine learning (ML) is the learning process developed for machines using various mathematical computational algorithms that can improve automatically through experience or by the use of data. It helps in visualizing various patterns to expect the potential data, or to execute crucial decision making under uncertain situations.

One of the most common algorithms that is used for predictive analysis in ML is **Linear Regression.**

In statistics, linear regression is a linear approach for forming a relationship between a scalar response to other independent and dependent variables. When there is a single input variable, the approach is termed as Simple Linear Regression, whereas when there are multiple input variables, it is termed as Multiple Linear Regression

**Gradient Descent** is a first order iterative optimization algorithm that finds a local minimum of a differentiable function. It is an algorithm that finds the best-fit line for a given training dataset in a smaller number of iterations.

**Linear Regression with Gradient Descent**

The aim of the learning process is to optimize the objective function. Gradient Descent is one of the supervised machine learning techniques that optimizes the cost function in the learning process.

The objective is to minimize the cost function and is defined as to understand the weights by using sum of squared errors between trained set and real outcomes.

The cost function defined is convex and powerful called gradient descent (incremental) to determine the minimum cost to classify the samples in the dataset. Partial derivation is computed on the cost function with respect to each weight.

Batch Gradient descent takes the entire batch as training set is a costly operation if slope is large. The incremental algorithm is preferred over batch gradient descent.

* 1. **Purpose of the Project**

Understanding the working of Linear Regression and its uses. Apply the concepts learnt in order to create a Linear Regression Model that predicts the sample outcomes of the dataset. Try to provide an optimal approach towards Linear Regression using Gradient Descent algorithm.

* 1. **Target Beneficiary**

It has the ability to help people across many domains like financial services or insurance domain, data analysts and users seeking results using predictive analysis.

* 1. **Project Scope**

Getting information from the web as datasets to run our project on and it includes gradient descent as optimization feature to make it more efficient version of linear regression

* 1. **References**

[1] Stochastic Gradient Descent using Linear Regression with Python, J V N Lakshmi, Research Scholar

Department of Computer Science and Application, SCSVMV University, Kanchipuram, India, 2016

[2] Linear Regression with Gradient Descent, Rukshan Pramoditha, 2020

1. **PROJECT DESCRIPTION**
   1. **Reference Algorithm:**

Gradient Descent Algorithm

* 1. **Characteristic of Data:**

Prediction of person's salary based on his years of experience.

* 1. **SWOT Analysis**

**Strength**

Strength is a characteristic that adds value to something by making it more special, unique and advantageous when compared. In this element of SWOT, the abilities and the key properties of organization are discussed that gives an organization an advantage over other organizations by making it more competitive. It defines the characteristics and situations of an organization which makes it more effective and efficient when compare with its competitors.

It defines the areas in which the organization hold a command or is good at doing it and that provides the organization and important capability. It can be a skill, a resource, image, market leadership, relation with buyer or supplier or any other advantage relative to its competitors that fulfil the needs of the market by providing the organization with a comparative advantage.

**Weakness**

Simple Linear Regression Weakness refer to the situation in which the existing capabilities and the resources the company holds are weaker or not sufficient compared to others organizations in the market. In other words, it means the aspects in which the organization is less efficient and needs to improve in order to align with the market trends. As these aspects negatively affect the overall performance of the organization by making it weaker compared to its competitors.

These are the factors that an organization lacks and does poorly in comparison to the organizations operating in the same market at the same level. It is a deficiency or limitation of resources, capabilities, skills that majorly affect the organizations effective performance. Management capabilities, Facilities, financial resources, marketing skills and the weak brand image can be the sources of weakness.

**Opportunities**

Simple Linear Regression Opportunity is an advantage and the driving force for an organization. It is the convenient time or situation that is present in the environment and will help the organization in achieving its goals. It is a factor that contribute positively towards the growth of the organization. It is a condition existing in the external environment that allow the organization to take an advantage of the organizational strengths, and help in overcoming the weaknesses and to neutralize the threats present in the environment.

**Threats**

Threats are the factors that prevent the organization from the actualization of an activity. It is an unfavourable situation that exist in the environment making it difficult for the organization to achieve its defined goals. It is a situation that arises as a result of the changes that took place in the immediate or distant environment, preventing the organization from maintaining its existence and superiority in the growing competition and are disadvantageous for the organization.

All the environmental factors are considered as a threat to an organization that could affect the efficiency and effectiveness of the organization.

* 1. **Project Features**

1. Reading and understanding the data
2. Data Preparation
3. Splitting the data into training and test sets
4. Building a linear model
5. Residual analysis of the train data

6: Making predictions using the final model and evaluation

* 1. **User Classes and Characteristics**

Identify the various user classes – general users

**2.6 Design and Implementation Constraints:**

Systems capable of executing c and cpp languages and installed with c and cpp libraries.

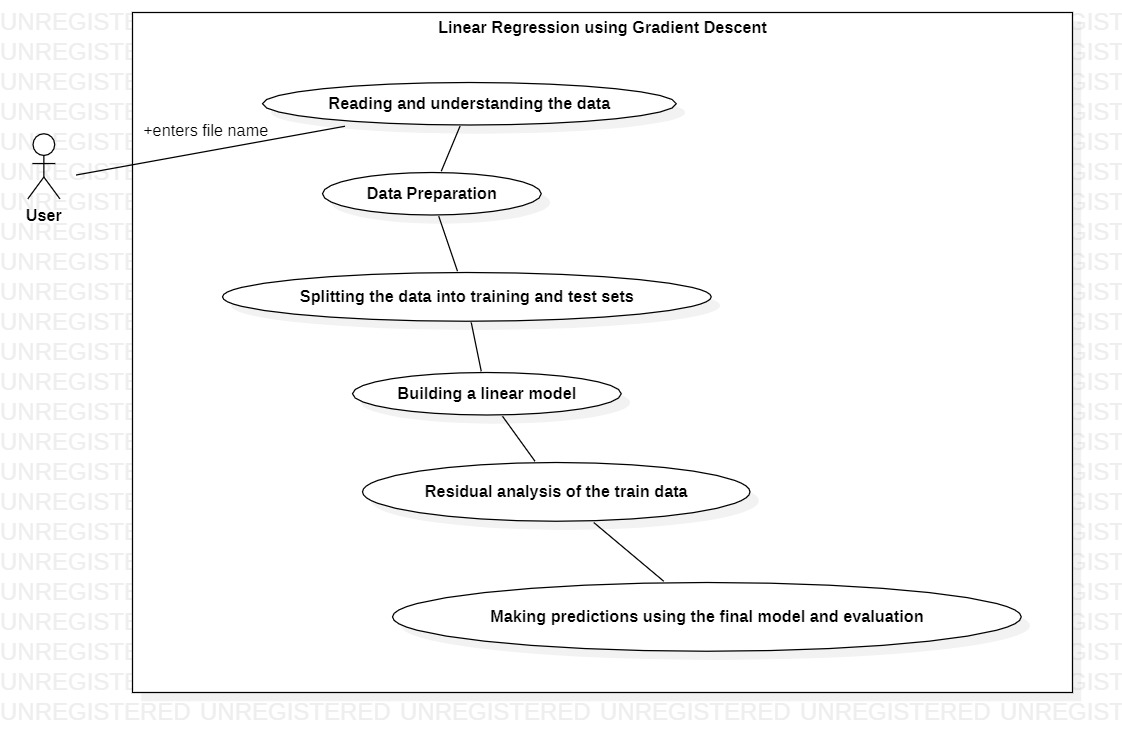
Systems installed with compiler MSV C++ 11.0 and higher.

Systems installed with basic cpp development IDE(windows and mac) or terminal(ubuntu and Linux) to execute the program.

**SOFTWARE REQUIREMENTS**

|  |  |  |
| --- | --- | --- |
| Name of Component |  | Specification |
| Operating System |  | Windows 10, Macintosh |
| Front end |  | C , C++ Programming Language |
| IDE Required |  | Visual Studio Code/XCode/codeblocks |
| **HARDWARE REQUIREMENTS** | | |
| Name of Component |  | Specification |
| Processor |  | Intel(R) Core (TM)i5-3210M CPY @ 2.50GHz 2.50 |
| RAM |  | 4GB |
| Hard Disk |  | 500GB HDD or 250GB SSD |
| Mouse |  | 2 or 3 Button mouse |
| Keyboard |  | 1. y Keyboard |

**2.7Design diagrams**



**2.8 Assumption and Dependencies**

Assumptions – Data is standardized so there is no need of standardization.

There are four assumptions associated with a linear regression model:

Linearity: The relationship between X and the mean of Y is linear.

Homoscedasticity: The variance of residual is the same for any value of X.

Independence: Observations are independent of each other.

Normality: For any fixed value of X, Y is normally distributed.

Dependencies – The project is dependent on the dataset.