*Major Project Report On*

***Machine Learning in Physics***

By

**Ammiel Peters (201900094)**

*In partial fulfilment of requirements for the award of the degree of*  
Bachelor of Technology in Computer Science and Engineering   
(2023)

Under the Project Guidance of

**Dr. Rakesh Moulick**

Scientific Officer-D

Centre of Plasma Physics-Institute for Plasma Research

And

Internal Reviewer

**Dr. Yumnam Nirmal**

Assistant Professor, SMIT

*DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING*



***(A constituent college of Sikkim Manipal University)***

**MAJITAR, RANGPO, EAST SIKKIM – 737136**

**PROJECT** **COMPLETION** **CERTIFICATE**

A close-up of a certificate

Description automatically generated with medium confidence

**i**

**Certificate of Acceptance**

This is to certify that the below mentioned student(s) of Computer Science & Engineering Department of Sikkim Manipal Institute of Technology (SMIT) has / have worked under the supervision of **Dr. Rakesh Moulick** from **23 Feb 2023 to 26 June 2023** on the project entitled “**Machine Learning in Physics**”

The project is hereby accepted by the Department of Computer Science & Engineering, SMIT in partial fulfillment of the requirements for the award of Bachelor of Technology in Computer Science and Engineering.

|  |  |  |
| --- | --- | --- |
| University Registration No | Name of Student(s) | Project Venue |
| 201900094 | Ammiel Peters | Centre of Plasma Physics – Institute for Plasma Research(CPP-IPR) |

Dr Udit Kumar Chakraborty

Professor & HOD

Computer Science & Engineering Department

Sikkim Manipal Institute of Technology

Majhitar, Sikkim – 737136

ii

**Declaration**

I, the undersigned, hereby declare that the work recorded in this project report entitled “**Machine Learning in Physics**” in partial fulfilment for the requirements of the award of the degree B.Tech (CSE) from Sikkim Manipal Institute of Technology (A constituent college of Sikkim Manipal University) is a faithful and bona fide project work carried out at “**Centre of Plasma Physics-Institute for Plasma Research, Nazirakhat, Sonapur, Kamrup(M), Assam**” under the supervision and guidance of  **Dr. Rakesh Moulick** of **Centre of Plasma Physics-Institute for Plasma Research.**

The results of the investigations reported in this project have so far not been reported for any other Degree / Diploma or any other technical forum.

The assistance and help received during the course of the investigation is duly acknowledged.

Ammiel Peters (Reg: 201900094)

iii

**Acknowledgement**

I wish to express my sincere thanks to External Supervisor **Dr. Rakesh Moulick, Scientific Officer D** of **Centre of Plasma Physics-Institute for Plasma Research** for providing me an opportunity to carry out project work on “Machine Learning in Physics” and their unlisted encouragement and guidance carrying out this project work.

I sincerely thank my **reviewer**/ **guide Dr. Yumnam Nirmal, Assistant Professor,** Computer Science and Engineering Department, Sikkim Manipal Institute of Technology for his guidance and encouragement in carrying out this project till the completion.

I would like to express my sincere thanks to **(Prof.)** **Dr.** **Udit Kumar Chakraborty**, HOD, Computer Science and Engineering Department for allowing me to carry out my project from **Centre of Plasma Physics-Institute for Plasma Research** and valuable support and guidance during the project period.

I would like to express my humble gratitude to **Mr. Biswaraj Sen, Professo**r, **Mr. Santanu Kumar Misra,** **Associate Professor, Mr. Saurav Paul, Assistant Professor-I, and Mrs Chitrapriya N., Assistant Professor-I,** Project Coordinators, Computer Science and Engineering Department, Sikkim Manipal Institute of Technology for their unlisted encouragement and their timely support and guidance till the completion of the project work.

Lastly, I wish to avail myself of this opportunity, express a sense of gratitude and love to all our teachers, staff of the department of Computer Science and Engineering Department, friends and fellow for their support and help.

.………………………….

Ammiel Peters (201900094)

**iv**

**DOCUMENT CONTROL SHEET**

|  |  |  |
| --- | --- | --- |
| 1 | Report No | CSE/Major Project/B.Tech/Internal/SA77/2023 |
| 2 | Title of the Report | Machine Learning in Physics |
| 3 | Type of Report | Research Project |
| 4 | Author | Ammiel Peters |
| 5 | Organizing Unit | Centre Of Plasma Physics – Institute For Plasma Research |
| 6 | Language of the Document | English |
| 7 | Abstract | In this project titled machine learning in physics we discuss the possible applications of machine learning in the field of space and laboratory plasma |
| 8 | Security Classification | General |
| 9 | Distribution Statement | General |

v

**LIST OF CONTENT**

**Chapter Title Page No.**

1. Abstract ix

1. Introduction 1

1.1 General Overview of Problem 1-2

1.2 Literature Survey 3-4

1.3 Problem Definition 5

1.2.1 Electron and the plasma interaction 5

phenomenon

1.2.2 Sunspot Data Analysis 5-7

1.2.3 Langmuir Probe 7-8

1.3 Proposed Solution Strategy 9

2. Implementation Details. 10-16

3. Results and Discussion 17-26

4. Summary and Conclusion 27

4.1 Summary 27

4.2 Conclusion 28

4.3 Limitations of the Project 29

4.4 Future Scope of Project 30

5. Gantt Chart 31

6. Reference 32

vi

**LIST OF FIGURES**

**Figure No. Figure Name Page No.**

1**.** Electron Plasma Phenomenon

2. Uneven graph (Predicted vs Actual Values)

3. Linear fit vs polynomial fit

4. Sunspot vs electron density at 500 Km

5. Sunspot vs electron density at 1000 Km

6. Sunspot vs electron density at 1500 Km

7. Sunspot vs electron density at 2000 Km

8. Sunspot vs electron temperature at 500 Km

9. Sunspot vs electron temperature at 1000 Km

10. Sunspot vs electron temperature at 1500 Km

11. Sunspot vs electron temperature at 2000 Km

12. 1st dataset with value 1

13. 2nd dataset with value 2

14. 3rd dataset with value 3

15. 4th dataset with value 4

16. 5th dataset with value 5

vii

**LIST OF TABLES**

**Table No. Table Name Page No.**

1. Electron-Plasma Data

2. Sunspot Data

3. Langmuir Probe Data

viii

**ABSTRACT**

Machine learning has become an important aspect of research in the field of physics, allowing for the extraction of valuable information from large and complex datasets. In this project report we discuss the possible applications of machine learning in the field of space and laboratory plasma. In the space arena, solar data has been collected and various relations have been established. Following this, the Langmuir probe data has been analysed for the laboratory plasma. Besides, this report also includes an overview of various applications of machine learning in physics, including but not limited to particle physics, astrophysics, and condensed matter physics. These applications include image and signal recognition, data classification, and prediction of physical properties. Additionally, some of the challenges associated with the use of machine learning in physics has been discussed, including the need for large amounts of data and overfitting. Despite these challenges, machine learning is proving to be a valuable tool for advancement of our understanding of the physical world, and is likely to play an important role in the field of physics in the years to come.

ix

**1. INTRODUCTION**

* 1. **General Overview of the Problems**

The main goals of this project are predicting the missing values from an experimental or numerical dataset and analysing them using machine learning algorithms. In this regard, we have used the following three types of experimental/numerical datasets:

* Problem-1: The beam electron and plasma interaction phenomenon in which we calculate the kinetic energy percent retrieval. The beam electrons are supposed to follow the magnetic field lines of the Earth and enter into the coexisting systems of hot and cold electrons, which are residual to the ionospheric atmosphere. The beam as and when appears in this upper Earth atmosphere, drives two stream instability. The beam usually loses the kinetic energy to the system of residual electrons, however, instantaneously it regains back a portion of the lost kinetic energy. This retrieval of kinetic energy again depends upon the initial drift speed of the beam. A pattern of such evolution is already available for certain values of the beam speed. We use the machine learning algorithm to find the appropriate fit of the given numerical data.

Diagram

Description automatically generated

Fig.1: The upper atmosphere of the Earth along with the magnetic field. lines

* Problem-2: The analysis of the sunspot data in terms of the electron density () and electron temperature () at various heights from the earth’s surface. Data has been
* taken from the year 2012 to 2022 for altitudes 500 km to 2000 km over Guwahati. The data has been collected from the website of the International Reference Ionosphere. In this case we have found the linear fits and appropriate correlation between parameters.
* Problem-3: Langmuir probe data has been analysed. Langmuir probe is a diagnostic tool for diagnosing plasma in the laboratory. Often such a probe is used by the spacecraft to diagnose the space plasma as well. It consists of a thin wire across which a sweeping voltage is applied to collect the plasma current. This gives rise to the standard I-V characteristics of a given plasma.

In this project many algorithms such as linear regression, decision trees, random forest, polynomial regression have been used for data prediction and analysis. Algorithms have been trained and tested on given datasets for outputs and then plotted for better visualization

**1.2 Literature Survey**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sl. No.** | **Author** | **Publication Details and Title** | **Findings** | **Relevance with work** |
| **1** | George Em KarniadakisIoannis G. KevrekidisLiu Yang [1] | Review Article 24  24 May 2021    Physics-informed machine learning | The rapidly developing field of physics-informed learning integrates data and mathematical models seamlessly, enabling accurate inference of realistic and high-dimensional multi-physics problems. | Got to know about machine learning algorithms in physics. |
| **2** | Iulia Georgescu | Feature  25 Jul 2022    How machines could teach physicists new scientific concepts | Artificial intelligence may uncover new scientific concepts that defy human intuition, but will researchers be able to understand and operate with them? This scenario might seem like science fiction, but physicists have faced it before. | Learnt how artificial intelligence have a great impact on physics. |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **3** | Yusuf H. Shaikh, A. R. Khan, M. I. Iqbal, S. H. Behere and S. P. Bagare | March 12, 2008  Sunspot Data analysis using time series | Analyzation of time series of sunspot data by R/S analysis and estimated the Hurst exponent H and the fractal dimension Df of the sunspot time series. | Understanding sunspot activity and predicting solar activity |
| **4** | K. O. Kiepenheuer | December 27, 1946  On the relations between ionosphere, sunspots and solar corona | Analysis of the ultra-violet solar radiation producing the ionospheric layers particularly its correlation to sunspots and coronal intensities. | We have got to know about the ionosphere layer from where we have collected the sunspot data. |
| **5** | Francis F. Chen | June 5, 2003  Langmuir Probe Diagnostics | Of all the ways to measure a plasma, the Langmuir probe is probably the simplest, since it consists of sticking a wire into the plasma and measuring the current to it at various applied voltages. However, it is an intrusive, not remote, technique; and the wire must be carefully designed so as not to interfere with the plasma nor be destroyed by it. | We have used all the datapoints through the Langmuir probe experiment which measures the current at various voltages. |

**1.3 Problem Definition**

**1.3.1 Plasma interaction phenomenon**

For the Problem-1 The given dataset contains the data of the normalized beam speed and the percent retrieval of the kinetic energy. It takes long time to do such a simulation and find out the individual data. This task is apparently tedious and time worthy. Therefore, the dataset may be trained on a computer and appropriate fit may be found out to predict any unknown data values. After predicting the missing values between the dataset we plot them using linear regression. While plotting the predicted values vs the actual values of the dataset between beam speed and retrieval percentage of kinetic energy we got an uneven graph

Table-1: The given dataset of the beam speed and the percent retrieval of kinetic energy

|  |  |
| --- | --- |
| **Normalized Beam Speed ()** | **% Retrieval of kinetic energy** |
| 15 | 23 |
| 20 | 30 |
| 25 | 25 |
| 35 | 10 |
| 40 | 20 |
| 45 | 40 |
| 60 | 85 |

**1.3.2 Sunspot Data Analysis**

In this part, we have analyzed the sunspot data. Variations of the electron density and temperatures have been noted for various altitudes from the Earth’s surface based on the data source (<https://irimodel.org/> and <https://www.swpc.noaa.gov/products/solar-cycle-progression>). Our attempt is to find the appropriate fit of the data along with finding the appropriate correlation between two data sets. The solar plasma is very dynamic in nature and over time the number of sunspots vary. Sunspots are the relatively cold region on the surface of the Sun. High sunspot number reflects high solar activity. In this case we compare the sunspot numbers with the electron

density and temperature at various heights so that we understand the effect of the solar activity on the upper atmosphere of the Sun. For the analysis, we have adopted a time ranging from 2012 to 2022 (month wise). Data corresponding to the year 2012 is shown in Table-2.

Table-2: Sunspot data for the year 2012.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Month\_Year** | **Sunspot Number** | **ne\_500** | | **Te\_500 K** | **ne\_1000** | | **Te\_1000 K** | **ne\_1500** | **Te\_1500 K** | **ne\_2000** | **Te\_2000 K** |
| Jan\_2012 | 94.4 | 1.47E+11 | 2801.2 | | 1.74E+10 | 3565.8 | | 6.64E+09 | 3882.5 | 3.78E+09 | 4281.7 |
| Feb\_2012 | 47.8 | 1.83E+11 | 2816.5 | | 1.74E+10 | 3458.7 | | 6.62E+09 | 3850.9 | 3.75E+09 | 4210.4 |
| Mar\_2012 | 86.6 | 2.35E+11 | 2733.2 | | 2.11E+10 | 3211.9 | | 8.01E+09 | 3832.2 | 4.54E+09 | 4179.4 |
| Apr\_2012 | 85.9 | 2.44E+11 | 2367.1 | | 2.62E+10 | 3083.7 | | 9.86E+09 | 3786.8 | 5.57E+09 | 4057.5 |
| May\_2012 | 96.5 | 2.13E+11 | 2019.5 | | 2.33E+10 | 3123.7 | | 9.94E+09 | 3680.7 | 5.59E+09 | 3792.6 |
| Jun\_2012 | 92 | 2.13E+11 | 1857.7 | | 2.33E+10 | 3220.1 | | 8.63E+09 | 3689.3 | 4.84E+09 | 3635.4 |
| Jul\_2012 | 100.1 | 1.76E+11 | 1856.1 | | 1.94E+10 | 3239.6 | | 7.21E+09 | 3698.2 | 4.04E+09 | 3623.1 |
| Aug\_2012 | 94.8 | 1.73E+11 | 2006.9 | | 1.95E+10 | 3284.1 | | 7.31E+09 | 3776.3 | 4.11E+09 | 3785.2 |
| Sep\_2012 | 93.7 | 1.91E+11 | 2245.9 | | 2.22E+10 | 3260 | | 8.44E+09 | 3733.7 | 4.79E+09 | 4046.3 |
| Oct\_2012 | 76.5 | 2.15E+11 | 2436.6 | | 2.56E+10 | 3243.7 | | 9.89E+09 | 3821.3 | 5.66E+09 | 4197.1 |
| Nov\_2012 | 87.6 | 1.98E+11 | 2605.1 | | 2.40E+10 | 3326.8 | | 9.34E+09 | 3908.5 | 5.36E+09 | 4083.9 |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Dec\_2012 | 56.8 | 1.71E+11 | 2749.3 | 2.03E+10 | 3463 | 7.84E+09 | 3908.2 | 4.49E+09 | 4087.4 |

**1.3.3 Langmuir Probe**

The Langmuir probe is a devise used for diagnosing the plasma in a laboratory setup. It gives the current versus voltage relationship of a plasma. From the data received by the probe, we are able to determine the electron temperature, electron density and ion density of the plasma. Our aim in this case is to find the best possible fit of the data using the available machine learning algorithms.

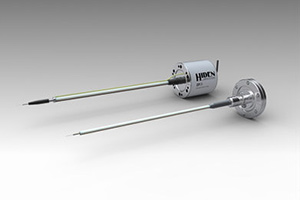


Fig: Langmuir probes (source: internet)

Table 3: Langmuir Probe Data

|  |  |  |
| --- | --- | --- |
| **Sl. No.** | **Voltage** | **Current** |
| 1 | -2.00E+02 | -4.17E-04 |
| 2 | -1.00E+02 | -4.16E-04 |
| 3 | -1.99E+02 | -4.16E-04 |
| 4 | -1.99E+02 | -4.15E-04 |
| 5 | -1.98E+02 | -4.14E-04 |

|  |  |  |
| --- | --- | --- |
| 6 | -1.97E+02 | -4.13E-04 |
| 7 | -1.97E+02 | -4.13E-04 |
| 8 | -1.96E+02 | -4.11E-04 |
| 9 | -1.95E+02 | -4.10E-04 |

**1.4 Proposed Solution Strategy**

* Python based Machine Learning tools such as Scikit learn along with standard python libraries have been used.
* We have applied decision tree to predict the missing values from the beam speed and the percent retrieval of kinetic energy dataset.
* Applying 6th degree polynomial regression in the electron and the plasma interaction phenomenon we get the best fit line.
* Applying linear regression to analyse the corelation in the sunspot dataset between sunspot and electron density and electron temperature.
* We have tried to applied polynomial regression on Langmuir probe dataset to analyse the flow of the data points.

**3. IMPLEMENTATION DETAILS**

As we have seen already the main aim of this project is to work on different types of physics related experimental datasets and do the analysis between them. All implementations have been done using python.

* **Electron and Plasma interaction phenomenon:** We have applied decision tree in the beam speed and the percent retrieval of kinetic energy dataset to predict the missing values and also plot the actual values with the predicted values.

#Pseudo Code

import pandas as pd

from sklearn.tree import DecisionTreeClassifier

from sklearn.model\_selection import train\_test\_split

import numpy as np

import matplotlib.pyplot as plt

# Open a file to write the results

file = open("file.txt", "w")

# Read the data from 'vd.csv' into a DataFrame

vd\_data = pd.read\_csv('vd.csv')

# Split the data into input features (X) and target variable (Y)

X = vd\_data.drop(columns=['value'])

Y = vd\_data['value']

# Split the data into training and testing sets

X\_train, X\_test, Y\_train, Y\_test = train\_test\_split(X, Y, test\_size=0.2)

# Create a DecisionTreeClassifier model

model = DecisionTreeClassifier()

# Fit the model to the training data

model.fit(X\_train.values, Y\_train)

# Ask for user input

n = int(input('Enter vd: '))

# Make a prediction for the input value

predictions = model.predict([[n]])

print(predictions)

# Write the predictions to the file for values from 0 to 99

for i in range(100):

    predictions = model.predict([[i]])

    file.write(str(i) + '\t' + str(predictions[0]) + '\n')

# Load the data from the file

x, y = np.loadtxt("file.txt", unpack=True)

T, Z = np.loadtxt("vd.txt", unpack=True)

# Plot the data

plt.plot(x, y, 'r.-', label='Predicted')

plt.plot(T, Z, 'b.-', label='Actual')

plt.grid(True)

plt.legend()

# Show the plot

plt.show()

Then we apply both linear and polynomial regression in this dataset and we observe that polynomial regression solves the problem which was occurring due to linear regression. We get the best fit line using polynomial regression.

# Pseudo Code

import matplotlib.pyplot as plt

import matplotlib as mp

import pandas as pd

import numpy as np

from sklearn.linear\_model import LinearRegression

from sklearn.preprocessing import PolynomialFeatures

# Load the data from 'ke\_analysis\_data.txt'

x, y = np.loadtxt('ke\_analysis\_data.txt', unpack=True)

# Reshape the x-array into a column matrix

x = x.reshape([len(x), 1])

# Create a LinearRegression model and fit it to the data

lin = LinearRegression()

lin.fit(x, y)

# Create PolynomialFeatures with degree 6 and transform the x data

poly = PolynomialFeatures(degree=6)

x\_poly = poly.fit\_transform(x)

# Fit the polynomial features and target values to the PolynomialRegression model

poly.fit(x\_poly, y)

lin2 = LinearRegression()

lin2.fit(x\_poly, y)

# Set matplotlib configurations

mp.rc('text', usetex=False)

mp.rc('font', family='sans-serif', size=10, serif='Computer Modern Roman')

mp.rc('axes', titlesize=10)

mp.rc('axes', labelsize=10)

mp.rc('xtick', labelsize=10)

mp.rc('ytick', labelsize=10)

mp.rc('legend', fontsize=10)

# Create a figure with two subplots

fig, ax = plt.subplots(1, 2, figsize=(10, 4))

# Plot the data and linear regression line in the first subplot

ax[0].plot(x, y, 'r.-', markersize=20)

ax[0].plot(x, lin.predict(x), color='blue')

ax[0].set\_title('Linear Regression')

ax[0].set\_xlabel('$v\_{d}$')

ax[0].set\_ylabel('$\Delta$ $KE\_{B}$(%)')

# Plot the data and polynomial regression line in the second subplot

ax[1].plot(x, y, 'r.-', markersize=20)

ax[1].plot(x, lin2.predict(poly.fit\_transform(x)), color='green')

ax[1].set\_title('Polynomial Regression')

ax[1].set\_xlabel('$v\_{d}$')

ax[1].set\_ylabel('$\Delta$ $KE\_{B}$(%)')

# Display the plot

plt.show()

# Predict a new result using Linear Regression

pred = 110.0

predarray = np.array([[pred]])

lin.predict(predarray)

# Predict a new result using Polynomial Regression

pred2 = 110.0

pred2array = np.array([[pred2]])

lin2.predict(poly.fit\_transform(pred2array))

**Sunspot Data Analysis:** A total of 133 data were collected from the international reference ionosphere over Guwahati city from the year 2012 to 2022 at distances 500, 1000, 1500 and 2000Km respectively. The analysis was between sunspot number vs electron density and electron temperature. Linear regression analysis has been used to analyse the data and get a appropriate corelation between them.

#Pseudo Code

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

# Set the figure size

plt.rcParams['figure.figsize'] = (20.0, 10.0)

# Read the data from 'Data.csv' into a DataFrame

data = pd.read\_csv('Data.csv')

# Extract the 'sunspot' and 'ne\_500m^-3' columns as arrays

X = data['sunspot'].values

Y = data['ne\_500m^-3'].values

# Calculate the mean of X and Y

mean\_x = np.mean(X)

mean\_y = np.mean(Y)

# Calculate the total number of values

m = len(X)

# Initialize the numerator and denominator for calculating the regression coefficients

numer = 0

denom = 0

# Calculate the numerator and denominator

for i in range(m):

    numer += (X[i] - mean\_x) \* (Y[i] - mean\_y)

    denom += (X[i] - mean\_x) \*\* 2

# Calculate the regression coefficients

b1 = numer / denom

b0 = mean\_y - (b1 \* mean\_x)

# Print the regression coefficients

print(b1, b0)

# Set the maximum and minimum values for x

max\_x = np.max(X) + 100

min\_x = np.min(X) - 100

# Generate evenly spaced values of x

x = np.linspace(min\_x, max\_x, 1000)

# Calculate the corresponding values of y based on the regression line equation

y = b0 + b1 \* x

# Plot the regression line

plt.plot(x, y, color='red', label='Regression Line')

# Plot the scatter points

plt.scatter(X, Y, c='green', label='Scatter Plot')

# Set the x and y axis labels

plt.xlabel('sunspot')

plt.ylabel('ne')

# Add a legend

plt.legend()

# Show the plot

plt.show()

**Langmuir Probe**: A Langmuir probe is a device that can determine the electron temperature, the electron density and the ion density of the plasma by obtaining its current-voltage characteristic since the characteristic curve depends on these plasma parameters. In this we have over 2000 experimental data divided into 5 datasets respectively. Polynomial regression is used to visualize the flow of the data points.

#Pseudo Code

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

from sklearn.linear\_model import LinearRegression

from sklearn.preprocessing import PolynomialFeatures

# Load the data for CH1\_Voltage and CH1\_Current

X = df['CH1\_Voltage'].values

Y = df['CH1\_Current'].values

# Reshape the X data into a column matrix

X = X.reshape(-1, 1)

# Create PolynomialFeatures with degree 7

poly = PolynomialFeatures(degree=7)

# Transform the X data using polynomial features

X\_poly = poly.fit\_transform(X)

# Fit the polynomial features and target values to the LinearRegression model

poly.fit(X\_poly, Y)

linreg = LinearRegression()

linreg.fit(X\_poly, Y)

# Predict the target values using the fitted model

Y\_pred = linreg.predict(X\_poly)

# Plot the scatter points

plt.scatter(X, Y, color='blue')

# Plot the predicted values

plt.plot(X, Y\_pred, color='red')

# Set the x and y axis labels

plt.xlabel('Voltage (V)')

plt.ylabel('Current (I)')

# Display the plot

plt.show()

**4. RESULTS AND DISCUSSIONS**

* When we applied decision tree on the beam speed and the percent retrieval of kinetic energy dataset. we got our missing values. But when we plotted the actual values vs the predicted(missing) values we got an uneven graph.

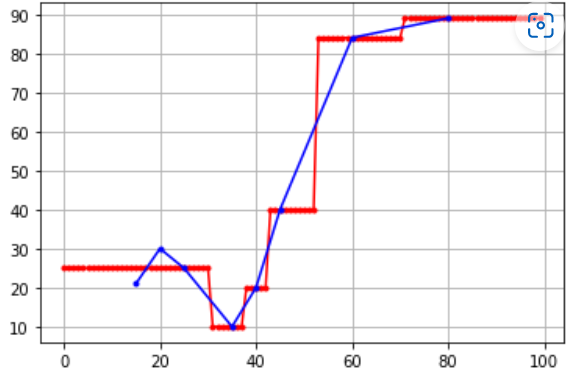


Fig.2

As we can see in the figure the blue lines are the actual points and the red line consists of predicted points and hence they does not respect each other.

**Electron-Plasma Interaction Phenomenon:** As we have applied both linear and polynomial regression in this phenomenon to check which one of those solves the problem.

In the figure below there is a detailed comparison linear fit and polynomial fit from which we can conclude that polynomial fit suits best for the problem.

Using the 7th order polynomial we have observed a perfect fitted graph.

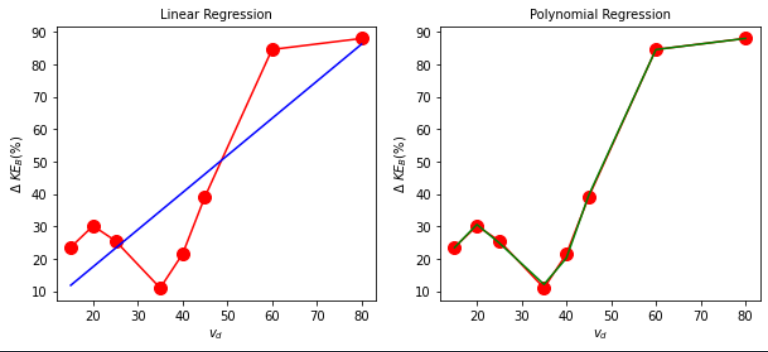


Fig.3

**Sunspot Data Analysis:**

* Firstly, we are analysing the sunspot number with the electron density at 500, 1000, 1500, 2000km.
* As we have seen the sunspot number vs electron density at 500 km gives us a positive linear curve.
* We have observed a positive correlation between sunspot number and electron density which is 0.8008

Chart, scatter chart

Description automatically generated

Fig.4

* At 1000 km we get a positive linear curve
* We have observed a positive correlation between sunspot number and electron density which is 0.8046

Chart, scatter chart

Description automatically generated

Fig.5

* At 1500 km we get a positive curve as well
* We got a positive correlation between sunspot number and electron density which is 0.8.

Chart, scatter chart

Description automatically generated

Fig.6

* Analysis at 2000km we get a positive linear curve.
* There is a positive correlation between sunspot number and electron density(ne^-3) which is 0.8043.

Chart, scatter chart

Description automatically generated

Fig.7

* Secondly we doing an analysis between sunspot number and electron temperature.
* Through observations we have observed that at 500 km we get a negative slope
* We have observed a negative correlation between sunspot number and electron temperature which is -0.046.

Chart, scatter chart

Description automatically generated

Fig.8

* At 1000 km we observe a positive slope
* At this height we got a positive correlation between sunspot number and electron temperature which is 0.5228.

Chart, scatter chart

Description automatically generated

Fig.9

* At 1500 km we get a positive curve
* We observe a positive correlation between sunspot number and electron temperature which is 0.65

Chart, scatter chart

Description automatically generated

Fig.10

* At 2000 km we observe a positive slope
* We have observed a positive correlation between sunspot number and electron temperature which is 0.4588

Chart, scatter chart

Description automatically generated

Fig.11

**Langmuir Probe:** Thereare five sets of entries in the given data set of the Langmuir probe. All the sets correspond to the current and voltage data received by the probe at five different time intervals for the same plasma parameters. Each dataset consists of 450 entries. Thus, there are total 2250 data points which are divided into 5 different datasets. Our developed routine is capable of extracting each data set from the given set of five data set and plot the results. The results are shown in the following figures. Using the polynomial fitting, we have observed that the seventh order polynomial gives the suitable fit for the data.

* 1st dataset with value 1: Applied polynomial regression with 7th order degree polynomial, we get a perfect fit line.

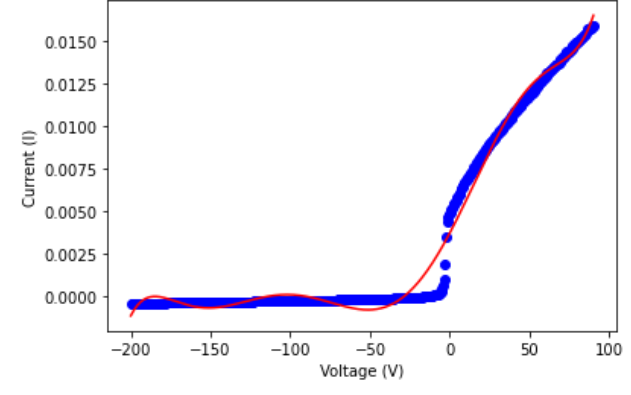


Fig.12

* 2nd dataset with value 2: Applied 7th degree polynomial.

A graph with a red line

Description automatically generated

Fig. 13

* 3rd dataset with value 3: On applying 7th order polynomial we observed

A picture containing text, screenshot, plot, line

Description automatically generated

Fig.14

* 4th dataset with value 4: We apply polynomial regression with 7th order polynomial

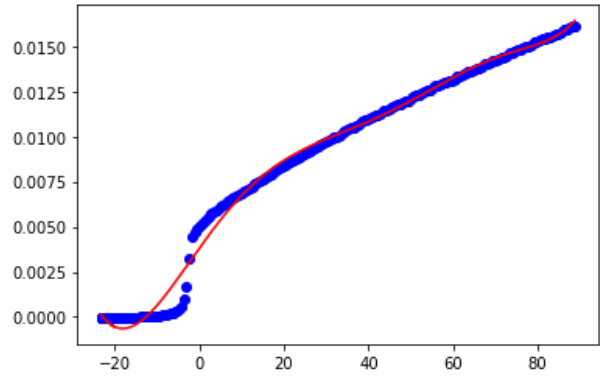


Fig.15

* 5th dataset with value 5: We apply polynomial regression with 7th order polynomial

A picture containing text, screenshot, plot, line

Description automatically generated

Fig.16

**5. Summary and Conclusion**

**5.1 Summary**

Machine learning has emerged as a powerful tool in physics research, offering new ways to analyse complex data, discover patterns, and make predictions. It has been successfully applied to various areas of physics, including particle physics, astrophysics, condensed matter physics, and quantum physics. Machine learning algorithms can efficiently handle large datasets, extract relevant features, and uncover hidden relationships that may not be apparent through traditional analytical methods. They have been used to solve problems such as particle identification, anomaly detection, image recognition, and data reconstruction. Additionally, machine learning techniques have enabled accelerated simulations, optimization of experimental setups, and the discovery of new physical phenomena.

The integration of machine learning with physics has opened up exciting possibilities for advancing scientific understanding and pushing the boundaries of knowledge.

In this research project we have used various experimental datasets and have done analysis using different types of machine learning algorithms.

**5.2 Conclusion**

Machine learning is revolutionizing the field of physics, offering innovative ways to analyse data, model complex systems, and make predictions. It complements traditional analytical techniques and provides a fresh perspective on challenging problems. By leveraging the power of machine learning algorithms, physicists can gain deeper insights into fundamental physical processes, accelerate scientific discoveries, and optimize experimental designs. However, it is essential to approach machine learning with caution and ensure proper validation and interpretation of results. The collaboration between physicists and machine learning experts is crucial for developing robust and reliable models that enhance our understanding of the universe.

As the field continues to evolve, the integration of machine learning techniques with physics research holds great promise for advancing our knowledge and addressing some of the most intriguing questions in the realm of physics.

**5.3 Limitations of the Project**

The major limitations faced during the project are:

* Machine learning algorithms heavily rely on high-quality and abundant data for training. However, in physics, obtaining large and reliable datasets can be challenging. Experimental measurements may have inherent uncertainties, limited resolution, or missing values.
* Many machine learning algorithms, such as deep neural networks, are often regarded as black boxes, making it challenging to interpret and understand how they arrive at their predictions.
* In physics, where explanations and causal relationships are highly valued, the lack of interpretability can be a significant drawback.
* Overfitting occurs when a machine learning model performs exceptionally well on the training data but fails to generalize to new, unseen dat

**5.4 Future Scope of the project**

This project can be published as research paper to understand the various uses of machine learning physics. Also, machine learning algorithms can help physicists analyse large datasets more efficiently and extract meaningful patterns or correlations. ML techniques can enhance the accuracy and speed of complex simulations in physics.

It can be also used to assist in optimizing experimental setups and parameters to achieve desired outcomes. By analysing historical data and patterns, machine learning algorithms can suggest optimal conditions for experiments, reducing trial and error and accelerating scientific discoveries.

**6. Gantt Chart**

A picture containing text, screenshot, number, font

Description automatically generated

**6. References**

[1] Karniadakis, George Em, et al. "Physics-informed machine learning." *Nature Reviews Physics* 3.6 (2021): 422-440.

[2] Georgescu, Iulia. "How machines could teach physicists new scientific concepts." *Nature Reviews Physics* 4.12 (2022): 736-738.

[3] Shaikh, Yusuf H., et al. "Sunspots data analysis using time series." *Fractals* 16.03 (2008): 259-265.

[4] Kiepenheuer, K. O. "On the relations between ionosphere, sunspots and solar corona." Monthly Notices of the Royal Astronomical Society 106.6 (1946): 515-524