IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

PREDICTION OF LAMINAR FLOW OF FLUID FLOW RATE THROUGH AN ORIFICE USING MULTI-LAYER PERCEPTRON

A. Mahesh Pramodh[1], M. Vamsi[2], V. Kranthi Kumar[3], Rasidha Bhanu[4], Dr T. Kiruba devi[5], Dr P. Dinesh Kumar[6]

[1,2,3] Undergraduate Student, Department of Computer science and Engineering, Dr.M.G.R Educational and Research Institute, Tamilnadu, India.

[4]Assistant Professor, Department of Computer science and Engineering, Dr.M.G.R Educational and Research Institute, Tamilnadu, India.

[5]Associate professor, Deputy H.O.D, Department of Computer science and Engineering (Artificial Intelligence), Dr.M.G.R Educational and Research Institute, Tamilnadu, India.

[6]Associate professor, Additional H.O.D, Department of Computer science and Engineering (Artificial Intelligence), Dr.M.G.R Educational and Research Institute, Tamilnadu, India.

ABSTRACT

This study delves into the comparative analysis of Linear Regression and Multi-Layer Perceptron (MLP) classifiers in the context of predicting flow rates. The focus lies on independent features such as temperature, pressure, square root of differential pressure, differential pressure, viscosity, and specific gravity, with the dependent feature being the flow rate 'Qv'. The Linear Regression model assumes a linear relationship between the independent and dependent variables, aiming to minimize the Root Squared Error (RSE). Meanwhile, the MLP classifier incorporates non-linearities through hidden layers, featuring 3 nodes in one layer and 2 nodes in another.

The Linear Regression model is represented by a linear equation, while the MLP classifier employs a multi-layered architecture for enhanced predictive capabilities. The study reveals that the MLP classifier achieves a significantly higher accuracy of 98% compared to Linear Regression. This higher accuracy underscores the MLP's ability to capture intricate patterns and non-linear relationships within the dataset, making it a more effective tool for flow rate predictions. The findings suggest that, in scenarios where complex relationships exist among variables, the MLP classifier outperforms traditional Linear Regression methods.

Keywords:

Flow rate predictions, Linear Regression, Multi-Layer Perceptron (MLP), Root Squared Error (RSE), Non-linear relationships.

1.INTRODUCTION

In this study, we delve into the intricate realm of predicting laminar fluid flow rates through an orifice, employing a sophisticated combination of Multilayer Perceptron (MLP) and regression techniques. The dataset hails from the renowned Iranian Oil Field, presenting a formidable set of 1038 features. Among these, the independent variables encompass crucial parameters like Temperature (T), Pressure (P), viscosity, differential pressure (ΔP), specific gravity, while the crux of our investigation lies in the dependent feature - the fluid flow rate (Qv).

Our pursuit revolves around harnessing the power of artificial intelligence, specifically MLP, to unravel the intricate relationships embedded within this extensive dataset. The MLP model, with its ability to comprehend complex patterns and non-linear dependencies, stands as a formidable tool for grasping the nuances of laminar fluid dynamics. Simultaneously, the incorporation of regression techniques adds a layer of interpretability, aiding in discerning the impact of each independent variable on the fluid flow rate.

The unique challenge posed by the Iranian Oil Field dataset necessitates a meticulous approach to feature selection, model training, and validation. Temperature, pressure differentials, viscosity, and specific gravity each play pivotal roles in determining laminar flow, requiring our models to decipher their collective influence accurately.

As we navigate through this computational journey, our primary aim is to develop a robust predictive model capable of estimating fluid flow rates with precision. The holistic understanding of these intricate interactions not only advances our comprehension of fluid dynamics but also holds immense practical significance for the oil and gas industry. This research amalgamates cutting-edge techniques with real-world data, offering a glimpse into the future of predictive modeling for complex fluid systems.

2.SIGNIFICANCE

The significance of flow rate in oil and gas fields is paramount, serving as a critical metric that underpins various operational aspects. Flow rate, often denoted as the volume of fluid passing through a specific point per unit of time, holds key importance in the oil and gas industry for several reasons.

Production Monitoring and Optimization

Flow rate is a pivotal parameter for monitoring and optimizing production processes in oil and gas fields. It provides real-time insights into the quantity of hydrocarbons extracted, allowing operators to adjust production strategies to meet demand and maximize resource recovery.

Reservoir Management

Understanding and controlling flow rates are essential for effective reservoir management. Accurate measurement of flow rates aids in evaluating reservoir performance, estimating reserves, and implementing strategies to enhance the recovery of oil and gas resources.

Equipment Sizing and Design

Properly sizing and designing infrastructure, such as pipelines and processing facilities, hinge on precise flow rate calculations. The capacity of equipment and pipelines is directly influenced by the anticipated flow rates, ensuring efficient and safe transportation of hydrocarbons.

Economic Analysis

Flow rate directly impacts the economic viability of oil and gas projects. Accurate assessments of flow rates are crucial for economic analyses, influencing investment decisions, project planning, and financial forecasting.

Health, Safety, and Environment (HSE) Considerations

Monitoring flow rates is essential for maintaining safe operating conditions. Deviations from expected flow rates can indicate potential issues such as leaks, blockages, or equipment malfunctions, helping to mitigate safety hazards and minimize environmental impact.

Regulatory Compliance

Many regulatory requirements in the oil and gas industry necessitate accurate measurement and reporting of flow rates. Compliance with these regulations is vital for maintaining operational licenses and ensuring responsible resource management.

3.TYPES OF FLOWS

Turbulent flow:

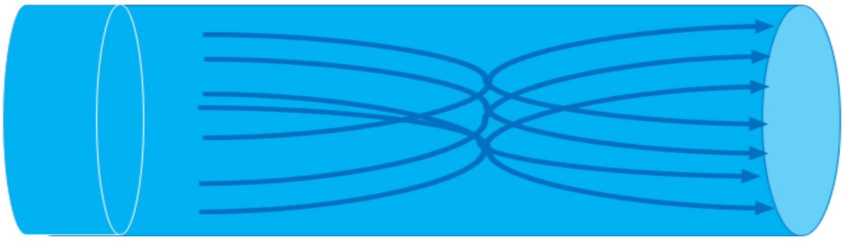


Fig. 1. *Direction of turbulent flow of fluid*

Turbulent flow is a fluid motion characterized by chaotic and irregular fluctuations in velocity, pressure, and other fluid properties [3]. In contrast to laminar flow, where fluid particles move in orderly layers, turbulent flow is marked by eddies, vortices, and swirls, creating a more complex and unpredictable pattern.

Darcy Weisbach equation:

The Darcy-Weisbach equation describes the relationship between flow rate, frictional head loss, and various parameters in a fluid system, commonly used for calculating flow rates in pipes. The equation is expressed as:

$$Q = \frac{K.A\sqrt{2gh_f}}{\sqrt{1-\left(\frac{D}{D_0}\right)^4}}$$

Whereas,

Q is the volumetric flow rate

k is a dimensionless constant (often 1 for S.l units)

A is the cross-sectional area of the flow

g is the acceleration due to gravity

$\mathbf{h_f}$ is the head loss due to friction

f is the Darcy-Weisbach friction factor

D is the diameter of the pipe

D_o is a reference diameter.

Laminar flow:

Fig. 2. *Direction of laminar flow of fluid*

Laminar flow refers to a smooth, orderly flow pattern in a fluid, typically characterized by parallel layers moving in the same direction [3]. This type of flow occurs at relatively low velocities and is in stark contrast to turbulent flow, which is chaotic and characterized by irregular fluctuations.

The fundamental governing equation for laminar flow in a pipe or channel is described by **Poiseuille's Law**. This law is derived from the Navier-Stokes equations under the assumption of steady, incompressible flow. For a Newtonian fluid, such as water or oil, Poiseuille's Law is expressed as:

$$Q = \frac{\pi r^4 \Delta P}{8\eta L}$$

Whereas,

Q is the volumetric flow rate

r is the radius of the pipe or channel

ΔP is the pressure drop along the length L of the pipe or channel

η is the dynamic viscosity of the fluid.

4.METHODS AND ALGORITHMS APPLIED AND COMPARED

MinMax Scaling method

MinMax scaling, also known as feature scaling, is a data preprocessing technique commonly used in machine learning to scale numerical features within a specific range. The method transforms the data, ensuring that it falls between a predetermined minimum and maximum value. This is particularly useful when dealing with algorithms sensitive to the scale of input features, such as gradient-based optimization algorithms.

The MinMax scaling formula is straightforward and can be expressed as:

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Here, **X** represents the original feature value, X_{min} is the minimum value of the feature in the dataset, and X_{max} is the maximum value. The resulting X_{scaled} is the rescaled value within the range [0, 1].

The process of MinMax scaling is applied to each feature independently. This ensures that all features contribute equally to the learning process, preventing dominance by features with larger scales.

Before MinMax Scaling

	Temperature	Pressure	Specific Gravity	ΔP	(ΔP)^0.5	Viscosity	qv
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	65.0	15.0	0.8950	11.34225	3.367826	22.31	31038.749558
2	22.0	18.0	0.8965	8.37225	2.893484	127.60	28742.518094
3	65.0	16.0	0.8975	8.83600	2.972541	22.31	27419.351077
4	70.0	13.0	0.8975	8.28100	2.877673	19.80	26476.381406
...	...	...	...	...	...	...	...
1033	97.0	65.0	0.8870	163.21600	12.775602	15.73	115467.655851
1034	96.0	67.0	0.8880	163.21600	12.775602	15.97	115439.798234
1035	97.0	65.0	0.8885	164.43025	12.823036	15.73	115799.997339
1036	97.0	67.0	0.8890	164.83600	12.838847	15.73	115911.442927
1037	97.0	70.0	0.8850	160.00000	12.649111	15.73	114456.626206

After MinMax Scaling

	Temperature	Pressure	ΔP	(ΔP)^0.5	Viscosity	qv
1	0.500000	0.056982	0.066068	0.244772	0.040527	0.246929
2	0.078431	0.069021	0.048693	0.207872	0.411253	0.227271
3	0.500000	0.060995	0.051406	0.214022	0.040527	0.215944
4	0.549020	0.048957	0.048159	0.206642	0.031689	0.207871
5	0.519608	0.040931	0.076411	0.264453	0.036759	0.265883
...	...	...	...	...	...	...
1033	0.813725	0.257624	0.954566	0.976630	0.017359	0.969716
1034	0.803922	0.265650	0.954566	0.976630	0.018204	0.969477
1035	0.813725	0.257624	0.961669	0.980320	0.017359	0.972561
1036	0.813725	0.265650	0.964043	0.981550	0.017359	0.973515
1037	0.813725	0.277689	0.935751	0.966790	0.017359	0.961060

Format method

The format method in Python is a powerful tool for formatting strings, enabling the customization of how data is displayed. In the context of converting continuous data into discrete data for classification, the format method can be used to control the precision of the numerical values, making them more suitable for classification tasks without explicit discretization.

Consider a scenario where you have a set of independent features and dependent features. To convert the continuous data into discrete data without discretization, you can leverage the format method to control decimal points. For instance, you might choose to display independent features with two decimal points and dependent features with zero decimal points.

	Temperature	Pressure	Specific Gravity	ΔP	(ΔP)^0.5	Viscosity	qv
1	0.50	0.06	0.90	0.07	0.24	0.04	0.0
2	0.08	0.07	0.90	0.05	0.21	0.41	0.0
3	0.50	0.06	0.90	0.05	0.21	0.04	0.0
4	0.55	0.05	0.90	0.05	0.21	0.03	0.0
5	0.52	0.04	0.90	0.08	0.26	0.04	0.0
...	...	...	...	...	...	...	...
1033	0.81	0.26	0.89	0.95	0.98	0.02	1.0
1034	0.80	0.27	0.89	0.95	0.98	0.02	1.0
1035	0.81	0.26	0.89	0.96	0.98	0.02	1.0
1036	0.81	0.27	0.89	0.96	0.98	0.02	1.0
1037	0.81	0.28	0.88	0.94	0.97	0.02	1.0

In this example, {:.2f} ensures that the independent features are displayed with two decimal points, while {:.0f} ensures that the dependent feature is displayed with zero decimal points.

This formatting approach retains the continuous nature of the data but adjusts the representation for classification purposes. Keep in mind that this is a visual representation and does not alter the underlying numerical values. It allows for a clearer display of the data while maintaining the integrity of the continuous features in a format suitable for classification algorithms.

Multi-Layer Perceptron

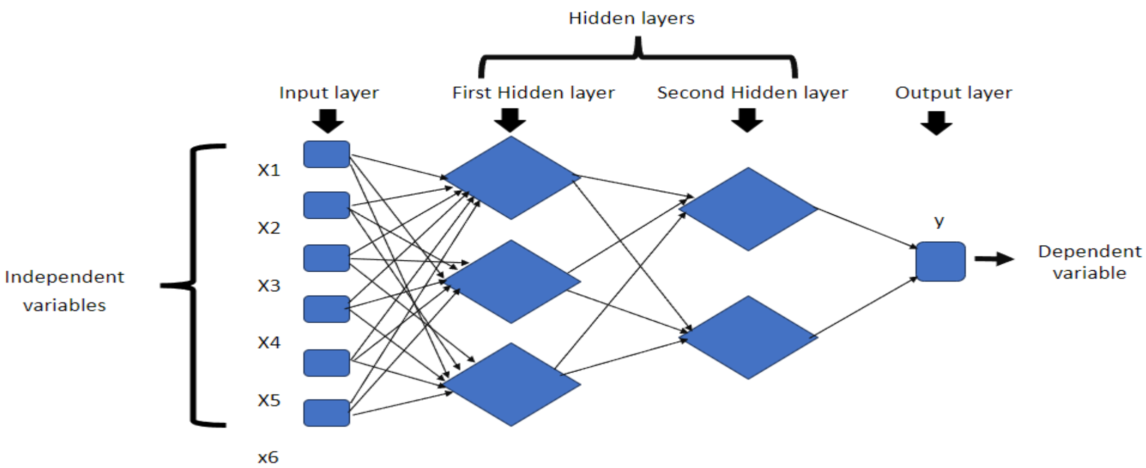


Fig. 3. *Architecture diagram of Multi-Layer Perceptron*

A Multi-Layer Perceptron (MLP) is a neural network architecture designed for complex tasks such as pattern recognition and classification. It comprises an input layer, one or more hidden layers, and an output layer [7]. Each layer consists of nodes (or neurons) interconnected by weighted edges.

The forward pass of an MLP involves the following steps:

1. Input Layer

The input layer receives the feature values (X_1, X_2, ..., X_n) of the input data.

2. Hidden Layers

Each node in a hidden layer computes a weighted sum of its inputs and passes it through an activation function. The weighted sum(z) is calculated as:

$$z_j = \sum_{i=1}^{n} w_{ij} x_i + b_j$$

Where w_{ij} represents the weight of the connection between input note ***i*** and output note ***j,*** and b_j is the bias for hidden note ***j.***

• The output of the hidden layer is obtained by applying an activation function (f) to the weighted sum:

$$a_j = f(z_j)$$

Common activation functions include sigmoid, hyperbolic tangent (tanh), or rectified linear unit (ReLU).

3. Output Layer:

The output layer performs a similar operation. The weighted sum for the output layer is computed as:

$$z_k = \sum_{j=1}^{m} w_{jk} a_j + b_k$$

where w_{jk} is the weight connecting hidden node ***j*** to output node ***k***, and b_k is the bias for output node ***k***.

• The final output is obtained by applying an activation function to the output layer:

$$y_k = f(z_k)$$

Training an MLP involves adjusting the weights and biases using backpropagation and optimization algorithms like gradient descent to minimize the error between predicted and actual outputs.

Correlation coefficients

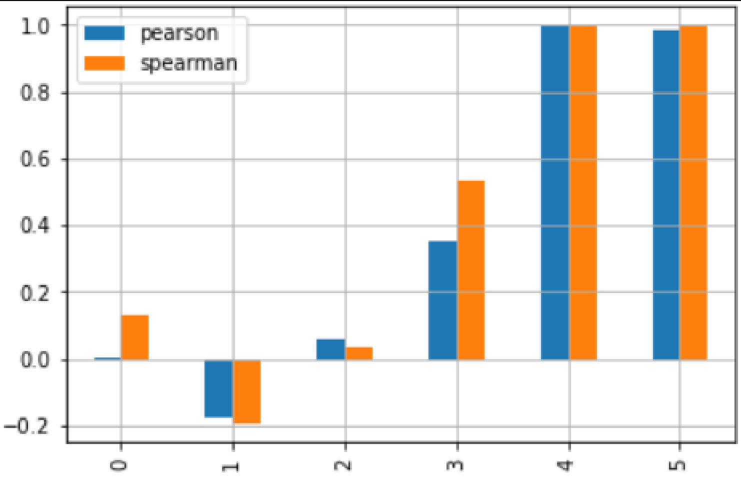
Correlation coefficients are valuable statistical measures used to quantify the strength and direction of relationships between variables. In your specific scenario, the correlation between the independent feature 'Δp' and the dependent feature 'qv' has been evaluated using both Pearson and Spearman correlation coefficients, resulting in higher values compared to other independent features.

Pearson Correlation Coefficient:

The Pearson correlation coefficient (ɼ) assesses linear relationships between two variables [9]. A coefficient close to 1 indicates a strong positive linear correlation. In your case, the Pearson correlation coefficient between 'Δp' and 'qv' is 0.987, signifying a highly positive linear association. This implies that as 'Δp' increases, 'qv' tends to increase proportionally, and vice versa.

Spearman Correlation Coefficient:

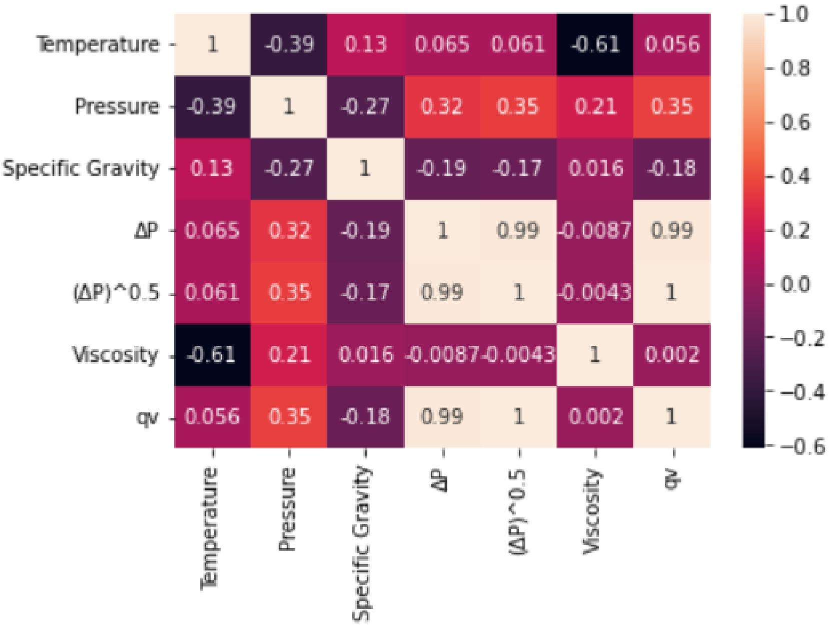
The Spearman correlation coefficient (ρ) measures monotonic relationships, capturing both linear and nonlinear associations [10]. A value close to 1 indicates a strong monotonic relationship. Here, the Spearman correlation coefficient between 'Δp' and 'qv' is 0.996, emphasizing a very strong monotonic association. This suggests that the variables not only have a linear correlation but also exhibit consistent monotonic behavior, reinforcing their connection.



Comparison with Other Independent Features:

The fact that 'Δp' has higher correlation coefficients (0.987 in Pearson and 0.996 in Spearman) compared to other independent features suggests that 'Δp' has a stronger and more consistent relationship with the dependent feature 'qv'. This can have implications in predictive modeling, as 'Δp' is likely a more influential predictor of 'qv' compared to other features.

The high correlation coefficients of 0.987 in Pearson and 0.996 in Spearman between 'Δp' and 'qv' indicate a robust and significant relationship. This emphasizes the importance of 'Δp' as a potential key factor in understanding and predicting the variation in the dependent feature 'qv' within the given dataset.



Linear Regression graph plotted between Δp and Qv.

A linear regression graph between the independent feature 'Δp' and the dependent feature 'Qv' visually represents the relationship between these two variables. In the graph, 'Δp' is plotted on the x-axis (horizontal), and 'Qv' is plotted on the y-axis (vertical).

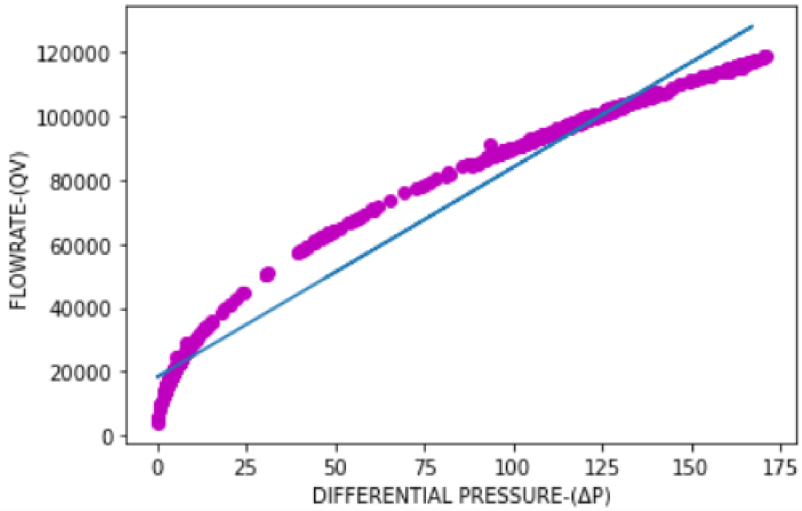
The graph typically exhibits a scatter plot of individual data points, where each point represents a specific observation of 'Δp' and its corresponding 'Qv' value. The linear regression line is then superimposed on the scatter plot.

The linear regression line is determined by the coefficients obtained during the training of the linear regression model. It represents the best-fit straight line that minimizes the sum of squared differences between the predicted 'Qv' values and the actual 'Qv' values for each 'Δp' observation.

If the linear relationship between 'Δp' and 'Qv' is strong, the data points should be relatively close to the regression line. The slope of the line indicates the strength and direction of the relationship, while the intercept represents the predicted 'Qv' value when 'Δp' is zero.

Interpreting the graph involves assessing the dispersion of data points around the regression line. A tight clustering of points suggests a strong linear relationship, while a more scattered distribution may indicate the presence of variability not captured by the linear model.

The linear regression graph visually encapsulates the relationship between 'Δp' and 'Qv', providing a clear depiction of how changes in 'Δp' correspond to changes in 'Qv' and facilitating a qualitative understanding of the linear regression model's predictive capability in this context.



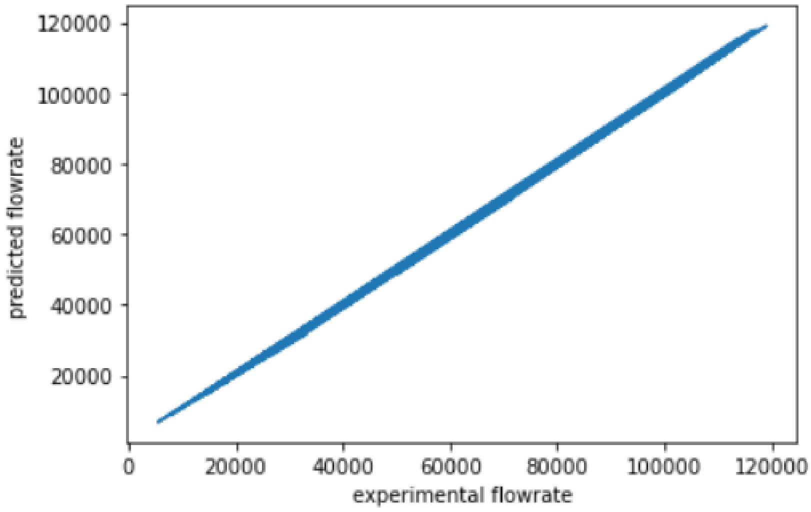
Linear Regression graph plotted between experimental and predicted flow rates post the removal of outliers.

The linear regression graph between experimental and predicted flow rates offers a visual assessment of the predictive accuracy of the model, particularly after the removal of outliers. On the graph, the x-axis typically represents the experimental flow rates, while the y-axis represents the corresponding predicted flow rates generated by the linear regression model.

The process begins with the removal of outliers from the dataset. Outliers, being data points significantly deviating from the general trend, can disproportionately impact the accuracy of the linear regression model. After outlier removal, the linear regression line is plotted over a scatter plot of the remaining data points.

The linear regression line represents the best-fit straight line that minimizes the sum of squared differences between the predicted and actual flow rates. If the model accurately captures the relationship between experimental and predicted values, the data points should align closely with the regression line.

The removal of outliers often results in a more robust and representative linear regression graph. Outliers, when present, can distort the slope and intercept of the regression line, leading to inaccurate predictions. The post-outlier removal graph provides a clearer picture of how well the model fits most of the data,



5.RESULT

Linear regression analysis and Multilayer Perceptron (MLP) classifiers serve different purposes in machine learning, with distinct characteristics and applications.

Linear Regression Analysis

Linear regression is a simple and interpretable model used for predicting a continuous target variable based on one or more independent variables. It assumes a linear relationship between the inputs and the output [2]. The model is trained by minimizing the residual sum of squares (RSS) or, equivalently, the root mean squared error (RMSE).

The linear regression analysis yielded a root squared error (RSE) of 0.96. However, linear regression might face limitations when dealing with complex, nonlinear relationships between features and the target variable.

Multilayer Perceptron (MLP) Classifier

An MLP classifier is a type of artificial neural network designed for classification tasks. It consists of an input layer, one or more hidden layers, and an output layer. Each layer has nodes with weighted connections, and nonlinear activation functions are applied to introduce nonlinearity into the model.

The MLP classifier with a configuration of 3 nodes in one hidden layer and 2 nodes in another achieved a higher accuracy of 98%. This suggests that the MLP classifier is better at capturing intricate patterns and relationships within the data compared to the linear regression model.

Accuracy

The MLP classifier demonstrated superior performance with 98% accuracy, indicating its ability to handle complex relationships in the data. In contrast, linear regression might struggle with capturing nonlinear patterns effectively.

Flexibility

MLP classifiers, with their multilayer architecture and nonlinear activation functions, are more flexible and capable of learning intricate representations. Linear regression, being linear, is less flexible in capturing complex relationships.

Interpretability

Linear regression models are generally more interpretable, as the relationship between input features and the target variable is expressed in a linear equation. MLP classifiers, with their hidden layers and complex interactions, are often seen as less interpretable.

6. CONCLUSION

While linear regression provides simplicity and interpretability, the MLP classifier, with its multilayer architecture, excels in capturing complex patterns, leading to a significantly higher accuracy in our specified scenario.

ACKNOWLEDGEMENTS

We would like to acknowledge the contributions of Dr.M.G.R Educational and Research Institute for providing access to necessary resources and facilities essential for conducting experiments and gathering data.

REFERENCES

1. Agwu, Okorie E. "Artificial Neural Network Model for Predicting Drill Cuttings Settling Velocity." *Petroleum*, vol. 6, no. 4, Dec. 2020, pp. 340–52, https://doi.org/https://doi.org/10.1016/j.petlm.2019.12.003.
2. Schneider, Astrid, et al. "Linear Regression Analysis." *Deutsches Ärzteblatt International*, vol. 107, no. 44, Nov. 2010, pp. 776–82. *PubMed Central*, https://doi.org/10.3238/arztebl.2010.0776.
3. Identifying the Types and Classifications of Fluid Flow Regimes. https://resources.system-analysis.cadence.com/blog/msa2022-identifying-the-types-and-classifications-of-fluid-flow-regimes. Accessed 12 Feb. 2024.
4. Ghorbani, Hamzeh. "Prediction of Oil Flow Rate through an Orifice Flow Meter: Artificial Intelligence Alternatives Compared." *Petroleum*, vol. 6, no. 4, Dec. 2020, pp. 404–14,https://doi.org/10.1016/j.petlm.2018.09.003.
5. Korkmaz, Murat Eray. "Using Artificial Neural Network for Predicting Performance of the Ranque–Hilsch Vortex Tube." *International Journal of Refrigeration*, vol. 35, no. 6, Sept. 2012, pp. 1690–96, https://doi.org///doi.org/10.1016/j.ijrefrig.2012.04.013
6. Kochkov, Dmitrii, et al. "Machine Learning–Accelerated Computational Fluid Dynamics." *Proceedings of the National Academy of Sciences*, vol. 118, no. 21, May 2021, p. e2101784118. https://doi.org/10.1073/pnas.2101784118
7. Sayahi, Tofigh, et al. "A RBF Model for Predicting the Pool Boiling Behavior of Nanofluids over a Horizontal Rod Heater." *International Journal of Thermal Sciences*, vol. 99, Jan. 2016, pp. 180–94. *ScienceDirect*, https://doi.org/10.1016/j.ijthermalsci.2015.08.010
8. E. Shirif, S.K. Elmabrouk, R. Mayorga, Application of Function Approximations to Reservoir Engineering, Lambert Academic Publishing, 2013, p. 432 13: 978- 3659280290.
9. Lin, Lawrence I.-Kuei. "A Concordance Correlation Coefficient to Evaluate Reproducibility." *Biometrics*, vol. 45, no. 1, 1989, pp. 255–68. *JSTOR*, https://doi.org/10.2307/2532051
10. L. Myers, M.J. Sirois, Spearman Correlation Coefficients, Differences between, Wiley Stats Ref: Statistics Reference Online, 2006.