International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4s
DOI: https://doi.org/10.17762/ijritcc.v11i4s.6536
Article Received: 19 January 2023 Revised: 22 February 2023 Accepted: 01 March 2023

Crime Prediction and Analysis against women Using LRSRI-Missing Value Imputation and FIPSO - Optimum Feature Selection Methods

P. Tamilarasi[1], Dr. R. Uma Rani[2]
[1]Research Scholar
Department of Computer Science
Sri Sarada College for Women(Autonomous)
Salem-636016, Tamil Nadu,India.
(ORCID: 0000-0001-9788-4129)
Email:tamilinresearch@gmail.com
[2]Principal
Sri Sarada College for Women (Autonomous)
Salem-636016, Tamil Nadu, India.
Email:principalsscsalem@gmail.com

Abstract—Data investigation is the method of considering crude measurements in arrange to draw conclusions around them. Many statistics evaluation techniques and tendencies had been automated into mechanical techniques and algorithms in such a manner that they provided raw statistics for human consumption. Machine learning could be a portion of artificial intelligence that permits computer frameworks to "analyze" their own statistics and improve them over time without being explicitly programmed. Machine learning algorithms can understand patterns in statistics and analyze them to make their own predictions. Lost esteem ascription is one of the foremost vital procedures in data pre-processing and it is additionally the most prepare of information examination. Ascription of lost information for a variable replaces lost information with a esteem inferred from an assess of the dispersion of that variable. Basic accusation employments as it were one suspicion. Numerous ascriptions employments diverse gauges to reflect the instability in evaluating this dispersion. In this article, The proposed method LRSRI used for impute the missing values on Crime against Women Data-set(CAW).The Linear Regression Imputation and Stochastic regression imputations are used in this method.Feature selection is another important data preprocessing techniques.This is often called attribute selection or feature selection. The most important problem in predictive modeling is the mechanical selection of features in the data. In this work,the proposed method FIPSO implemented for feature selection.This is feature importance and Particle Swarm Optimization based method.The main objective of this work is predict the crime rate against women in India based on 2001 to 2021 crime recorded against women in India.This Data set is collected from Data.gov.in.Finally The predicted result is compared with recent NCRB crime report.The proposed method LRSRI and FIPSO has given 98.34% accuracy of crime prediction.In feature,This outcome will be valuable for the crime office to control the CAW in India

Keywords- Crime Prediction, Particle swarm optimization, Missing Value Imputation, Feature Selection.

I. INTRODUCTION

Data analysis is recognized as an vital unused logical teach and prime example driving inquire about and advancement in regions such as insights, computer science, and insights science, and is broadly utilized in science, designing, the open division, financial matters, social sciences, and lifestyle.We are driving practical change in areas such as: Data analysis involves the extraction and classification of data to identify and analyze behavioral data and patterns, and how this is done depends on the needs of the organization. Data analysis is also called "data analysis." Data analysis is a procedure of investigation, pre-processing, transforming, and modeling to find useful information and draw conclusions for valuable decision-making. Data pre-processing is a technique of data analysis. Modify raw information into a readable format. Real data is always incomplete, inconsistent, lacks specific behavior or trends, and can contain many errors. Information preprocessing could be a recognized strategy for tackling comparative authoritative issues. Ascription is an approach to supplant lost input with surrogate values to protect most of the information data in a information set. These procedures are used because it is not possible to remove data from the data-set each time and the size of the data-set can be significantly reduced, raising concerns about data-set bias as well as errors. because it can also lead to causes of analysis. Use insinuation because missing data can beget the following problems

- ✓ When using libraries for ML (the most popular being skLearn), there is no automatic way to handle this

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4s
DOI: https://doi.org/10.17762/ijritcc.v11i4s.6536
***Article Received**: 19 January 2023 **Revised**: 22 February 2023 **Accepted**: 01 March 2023*

missing data, which can point to errors.

- ✓ Large amounts of missing data can skew the distribution of variables. it can increase or decrease the values of specific categories in data set.
- ✓ Missing data can skew the dataset and lead to incorrect analysis by the model.

A regression is just a technique that is statistical relates a reliant variable to one or higher independent variables. A regression model has the capacity to show whether changes observed in the reliant adjustable are related to alterations in one or more regarding the factors which can be explanatory.It does this by basically fitting a line that is best-fit seeing the way the information is dispersed surrounding this line. Regression helps economists and economic analysts in things which range from asset valuation to predictions which can be making. To ensure that regression leads to be properly interpreted, a few presumptions concerning the information additionally the model it self must hold.In this work, LRSRI methods is proposed to impute the missing values.This is a hybrid method of Linear and Stochastic linear regression techniques.

Many attributes will be produced when a machine learning model is built using a real data-set, but not all of these attributes will always be crucial. Including superfluous highlights amid demonstrate preparing decreases the model's in general precision, complicates the demonstrate, and diminishes the model's capacity to generalize and inclination the model. Attribute choice is one of the foremost vital steps in making a machine-learning model. Common machine learning feature selection techniques include:

- ✓ Filter method
- ✓ wrapper method
- ✓ embedded method

Filter methods are typically used in pre-treatment steps. These methods, apart from using machine learning algorithms, select features from a data-set. Computationally, they are very fast and cheap, and are very good at removing duplicate, correlated, and redundant features, but these methods do not remove multicollinearity. A wrapper method, also called a greedy algorithm, iteratively uses a subset of features to train the algorithm. Features are added and removed based on conclusions drawn from training in front of the model. The stopping criteria for choosing the best subset are usually predefined by the person training the model. When the model's performance degrades or reaches a certain number of characteristics. The main advantage of wrapper methods over filter methods is that they provide an optimal set of functions for training a model, so they are more precise than filter strategies but more computationally costly. The inherent method has its own built-in feature selection method because the feature selection algorithm is integrated as part of the learning algorithm. Embedding methods counter the discriminate of filter and wrapper methods and combine their advantages. Like filtering methods, these methods are faster and more accurate than filtering methods, and feature combinations are also considered.

The concept of optimization is essential in getting to know a device. Most devices that get to know fashions use education information to study the connection between input and output information. This version may be used to make approximate predictions about developments or classify new information. This is an optimization procedure aimed at enhancing the model's accuracy and lowering the mistake rate with every iteration. Optimization is a content that runs through every expedient of machine literacy. This includes a data scientist optimizing and enriching labeled training data or the iterative training and enhancement of models. At its center, planning a machine instruction illustrate is an optimization issue, as the appear learns to perform a work most effectively. The first basic parcel of machine capability optimization is squeezing and standardization appear courses of action or hyperparameters.The problem of crimes against women is a global problem. One in three women worldwide will witness physical or sexual violence in her continuance. This implies that approximately 1 billion ladies are influenced by this point. Assault, residential viciousness, sexual attack, and kill are the foremost common viciousness against ladies. Most violent acts are motivated by implied retribution, dowry issues, and forced prostitution. Violence against women can affect anyone, regardless of caste, religion, creed, or race

There are several factors that contribute to the high crime rate against women. These include issues of poverty, gender inequality and ineffective law enforcement. In multitudinous countries, women do not have the same legal rights as men and are constantly treated as alternate- class citizens. The issue of crime against women is complex, but it's important to address it to cover women from detriment.This gets them more undefended to abuse and development. Poverty also increases the risk of violence against women, as desperate people may turn to crime to survive. Ultimately, weak law enforcement means that perpetrators often get away with their crimes, leaving victims feeling helpless and without justice. The issue of crime against women is complex, but it is important to address it to protect women from harm. In this article,FIPSO method proposed for select the features on CAW datasets.This hybrid techniques for Feature Importance and Particle Swarm Optimization methods.This dataset holds 325 rows and 24 columns.

II. RELATED WORKS

The author **Napsu Karmitsa et.al[1]** introduced a technique that is new for incomplete data . The technique is

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4s
DOI: https://doi.org/10.17762/ijritcc.v11i4s.6536
***Article Received**: 19 January 2023 **Revised**: 22 February 2023 **Accepted**: 01 March 2023*

dependant on cluster linear that is wise and it combines two well-known approaches for missing value imputation: linear regression and clustering. The proposed proficiency is proven using some unreal and real-world datasets and unlike other rule due to missing value imputations. Numerical results show that this method is likely to provide the most accurate calculations for the MCAR and MAR datasets, as its structure shows a percentage of missing data up to 25%.The researcher **Azza Ali et.al[2]** proposed fuzzy Matching that is k-Top Value(FKTM) for missing value imputation. It imputes lacking numerical and categorical content with smart estimates according to comparable records, decreasing bias. Expectation-maximization is employed, where it employs clustering that is fuzzy find a group of comparable data and estimate them.The proposed approach outperforms MICE with an accuracy of 82.2. The proposed approach outperforms techniques with an accuracy of 86.6 % on the cryotherapy dataset.The author **Mohammed.H et.al[3]**presented a hybrid model imputation method (HIMP) developed to deal with various missing data patterns in the real-time IRDia dataset. HIMP belongs six steps: missing data analysis, missing data imputation using MNAR patterns, disintegration, single accusation, multiple accusation methods.The proposed method accuracy were compared with other classifier algorithms accuracy.Finally the proposed algorithms more efficient than others.The author **Rahin Atiq et.al[4]** implemented four imputation that is different such as for instance regular value, suggest, KNN, MICE to replace the values which can be missing utilize them to produce prediction models and applied six various types of classifier techniques to prediction of voucher offers for customers.The deep learning classifier get better accuracy than other classifiers.The researcher **Amelia Ritahani Ismail et.al[5]**provides a review that is comprehensive of various imputation techniques used to replace the lacking information. Purpose of the assessment paper is to bring attention that is particular to improvements that are prospective existing methods and supply visitors with a better grasps of imputation method trends.**Muhammad Metwally Seliem[6]** suggested TOMI method for handling outliers data as missing data and applied various L algorithms for prediction.finall the author concluded that NB and LR algorithms prediction accuracy is outperformance compare with all other algorithms.The writer **Saba Bashir[7]** proposed a novel framework is proposed which utilizes attribute that is significantly diffent methods from filters, wrappers, and embedded algorithms. Also, classification is at that point performed on chosen highlights to classify the information employing a back vector machine classifier. Two benchmark that is publically available are utilized, i.e., the Microarray dataset as well as the Cleveland Cardiovascular sicknesses dataset, for experimentation and examination, furthermore they are chronicled from the UCI data store. The execution of SVM is analyzed precision that's utilizing affectability, specificity, and f-measure. The exactness of 94.45% and 91% is gotten on all dataset, individually.The researcher **Kaushalya Dissanayake[8]** has done an experimental assessment of the performance of models produced category that is utilizing and appropriate features selected using function that is significantly diffent approaches. For comes about of this examination that's exploratory ten work determination procedures, i.e., ANOVA, Chi-square, shared data, ReliefF, forward work choice, backward function selection, exhaustive function selection, recursive feature reduction, Lasso regression, and Ridge regression, and six category approaches, in other words., decision tree, random forest, help vector machine, K-nearest neighbor, logistic regression, and Gaussian naive Bayes, take place placed on Cleveland cardiovascular illnesses dataset. The function subset selected by the big event that is backward strategy has accomplished the category precision that is greatest of 88.52%, precision of 91.30per cent, sensitiveness of 80.76%, and f-measure of 85.71% aided by the choice tree classifier.The researcher **Tarneem Elemam[9]** proposed a ML algorithm proposed to diagnose cancer tumors that differs from big information. The algorithm comprises a two-stage feature selection that is hybrid. A general ranker is initiated to combine the results of three filter-based function evaluation techniques, particularly, chi-squared, F-statistic, and mutual information (MI) in the first stage. The features are then ordered according to this combination. The modified wrapper-based sequential forward selection is employed to find the optimal feature subset, using ML models such as support vector device (SVM), decision tree (DT), random woodland (RF), and K-nearest neighbor (KNN) classifiers in the 2nd stage. To look at the proposed algorithm, many tests have now been carried out on four micro array that is malignant, employing in the process 10-fold cross-validation and hyper-parameter tuning. The performance associated with the algorithm is examined by calculating the precision that is diagnostic. The outcome suggest that for the leukemia dataset, both SVM and KNN models register the accuracy that is highest at 100% only using 5 features. For the ovarian cancer dataset, the SVM model achieves the accuracy that is greatest at 100% using only 6 property.**Namjung Kim[10]**observed 45 and 249 characteristics that can viably be fundamental 896 zeolite descriptors made by the matminer system for calculating the majority and deformation moduli of zeolites, severally. A database containing the specialistic properties of 873 zeolite construction, deliberate thickness that's making utilize of speculation, was utilized to taught the hardware relapse appear that's learning. The results of using these features which are often critical the LightGBM algorithm were rigorously weighed against those from other

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4s
DOI: https://doi.org/10.17762/ijritcc.v11i4s.6536
***Article Received**: 19 January 2023 **Revised**: 22 February 2023 **Accepted**: 01 March 2023*

regressors along with s with different sets of features. The contrast results suggest that the model that is surrogate critical features advances the forecast accuracy for the majority and shear moduli of zeolites by 17.3per cent and 10.6%, respectively, and decreases the forecast doubt by one-third with this accomplished using features that are previously available.

III. PROPOSED METHODOLOGY

3.1 LRSRI - Missing Value Imputation

Linear relapse could be a gadget that's directed demonstrate majorly found in estimating. Checked machine learning models are those where we utilize the preparing data to construct the demonstrate and test the precision at that point with respect to the model using the misfortune work. Straight relapse is one of the foremost broadly known time appear determining procedures that's used for prescient modeling. Due to the truth title suggests, it expect a relationship that's direct a few of autonomous components to that for the subordinate adjustable.

In this article, the lost values are ascribed by utilizing direct relapse and stochastic relapse.

✓ Linear Regression Imputation:

The measure that is missing changed for the predicted value created by the regression of the lacking product on items observed for the system. Mean imputation are seen as a situation that is unique of imputation in which the predictor variables are dummy indicator factors for the cells within that the means are imputed.Below equation used for linear regression.

$$Y = mx + c \quad (1)$$

Where y signified the subordinate variable

M alludes to the slant with respect to the line

X is the free which is adjustable

C is for steady

✓ Stochastic relapse ascription started to be able to resolve this presssing predicament of deterministic relapse ascription. Stochastic relapse ascription includes a botch that's arbitrary to the anticipated esteem and is subsequently in a position to duplicate the relationship of X and Y more properly.The underneath fig 1 appears the work stream of lost values ascription. The underneath equation 2 is utilized for stochastic relapse

$$y \; x \; e = \beta + \beta + -\beta \; \eta \quad (2)$$

Here y x e is random.

β is unobserved variable.

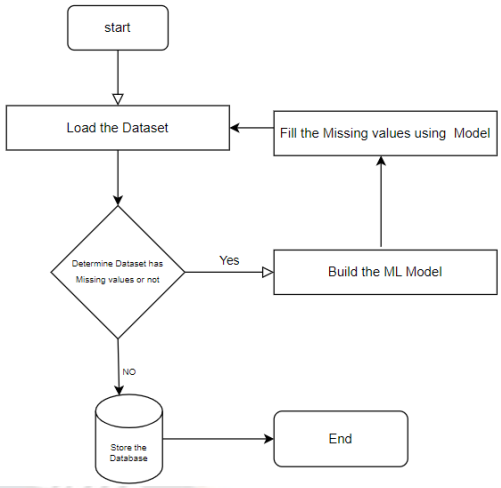

Fig 1: Missing Values Imputation Work Flow

Missing Values Imputation Algorithms using LRSRI Methods.

Algorithm _1 : Linear and Stochastic regression used to fill the Missing data.
Input : Missing value in the column (MVA) in Crime_Dataset(CAW). Output : Complete the missing field with regression Value (LRSR) for all field (F).{f1,f2,f3. fn }.
Step 1: Read the CAW data-set along with property {f1, f2, f3.....} Step 2: Find the null values Step 3: Separate the blank field to test data. Step 4: Drop the empty cell and assigned it to the train element Step 5: Make x_train and Y_train data from the train field Step 6: Build the Model i.e linear and stochastic Step 7: Make x_test from test. Step 8: Apply X_test for Prediction Step 9:Replace the blank field with the help of Predicted values.

3.2 FIPSO -Feature Selection Methods

✓ Extra Tree Classifier

Typically a model-based approach that employments a tree-based administered show to choose highlights and decide include significance. Additional Tree Classifier or Amazingly Irregular Tree Classifier is an gathering calculation that triggers numerous randomly constructed tree models from a preparing information set and disposes of the foremost voted highlights. Rather than bootstrapping reproductions, we fit each choice tree to the complete dataset and arbitrarily select part focuses to part the nodes.The part of hubs that happens at each level of the setup tree is based on the degree of

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4s
DOI: https://doi.org/10.17762/ijritcc.v11i4s.6536
***Article Received**: 19 January 2023 **Revised**: 22 February 2023 **Accepted**: 01 March 2023*

arbitrariness or randomness of the subnodes. Hubs are part over all factors accessible within the information set and the part driving to the foremost homogeneous children is chosen within the development tree demonstrate. This decreases change and makes the demonstrate little inclined to overfitting. The underneath figure 2 appears the Highlight determination utilizing proposed strategy FIPSO.

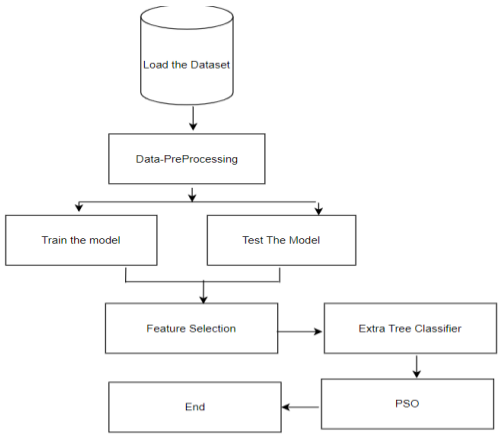

Fig 2: Work Flow of Feature Selection

✓ PSO (Particle Swarm Optimization)

PSO is change of state that's stochastic proposed by Kennedy and Eberhart (1995) . A populace based look scheme with perspective of particle is grouping that's utterance to Swarm of particles as superficial generalist. PSO is antitrust a strong optimization that's biological process on the premise of the development and insights of swarms. PSO insight the minimum esteem for the enormous occasion. The statement "PSO" is incomparable to a run of fowls trenchant for nourishment.bird = particle,food = solution

pbest = best response the particle has ever achieved.

Gbest = best global solution for all particles in the swarm.The basic approach of PSO is individual acceleration.Particles going to the pbest and gbest neighborhoods randomly adequate acceleration at any point in time.PSO uses many agents. In other words, the particles form a swarm that flies around the search space in search of the optimal solution.Each tick can be thought of as a point in an N-dimensional domain that coordinates its 'flight' parallel to its own contribution and that of another particle. Each particle attempts to change its position P using the below formula formula 3 and 4.

$$P(t+1) = P(t) + w(t+1) \quad (3)$$

$$w(t+1) = w1w(t) + c1 \times \text{rand}() \times (\text{Xybest} - P(t)) + c2 \times \text{rand}() \times (\text{Pgbest} - P(t)) \quad (4)$$

Here,

w(t) -> velocity of the particle at time t

P(t) ->Particle position at time t

W1->Inertia weight

c1 , *c2* ->learning factor or accelerating factor

Rand -> uniformly distributed random number between 0 and 1

Ppbest->particle's best position

*Pgbes*t->global best position

The below algorithm 2 give details of feature selection of CAW using proposed method FIPSO

Algorithm _2 : Feature Importance and PSO techniques used to select the features. Input : Load the CAW dataset with 22 features Output: Select the important features for crime prediction using FIPSO method
Step 1: Read the CAW dataset with various states and crime types with the complete dataset. Step 2: Select the features using FI Extra Tree Classifier method Step 3: Evaluate the features Step 4: If the evaluation is good Step 5: Stop the criterian and validate the model Step 6: else Step 7: Select the features using PSO techniques Step 8: continue the Step3,Step 4,Step 5 Step 9:End the process

IV. Results and Discussion

The below figure 3 shows cruelty is the highest crime type in India from 2016 to 2022.

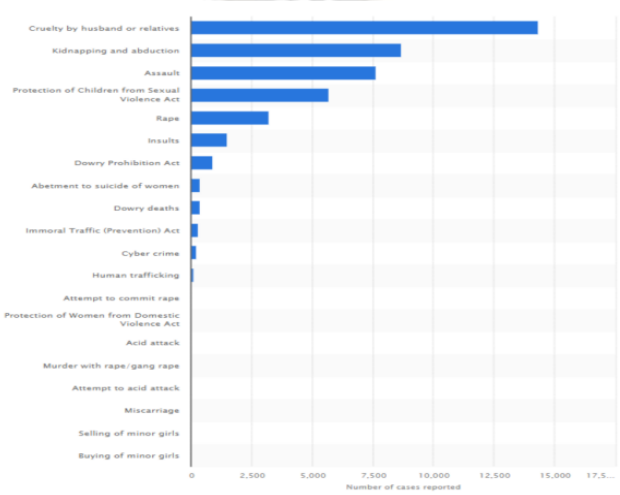

Fig 3: Highest Crime categories in India.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4s
DOI: https://doi.org/10.17762/ijritcc.v11i4s.6536
***Article Received**: 19 January 2023 **Revised**: 22 February 2023 **Accepted**: 01 March 2023*

According to the NCRB(National Crime Recorded Bureau)report the maximum number of cruelty crime type recorded in 2022. It showed in figure 4.

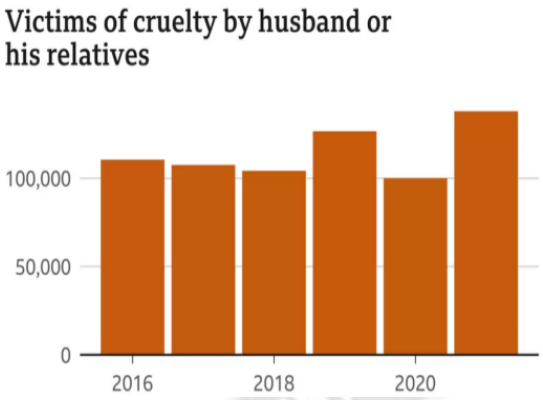

Fig 4:Highest cruelty crime tpe recorded in 2022

The below fig 5 shows various states crime registered details in,India from 2016 to 2021.Here, UtterPradesh is the highest crime recorded state compare than other state.

Fig 5: Highest Crime rate state in India.

The below fig 6 shows the missing values in every features.

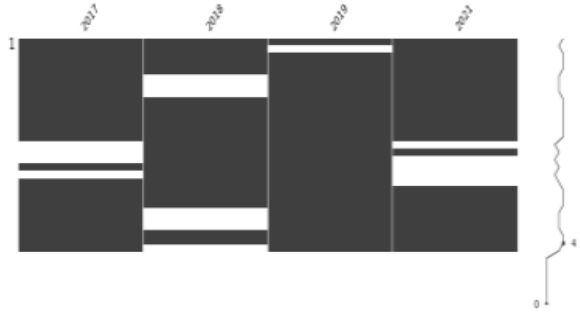

Fig 6: Crime against Women Data-Set with Missing Values.

The following fig 7 shows the complete dataset with the help of proposed algorithms LRSRI algorithm.

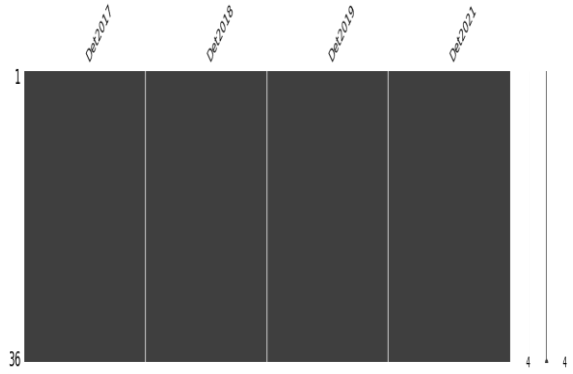

Fig 7:After filling the missing values CAW data-set

```
# display the relative importance of each attribute
print(model.feature_importances_)
```

```
[0.10527809 0.09669154 0.09781526 0.08827156 0.09092312 0.10229624
 0.11013963 0.10409632 0.10208611 0.10240211]
```

Fig 8:Select the features using Extra Tree Classifier

Figure 8 above shows feature selection using the Extra Tree Classifier.The below figure 9 shows feature selection using PSO methods.Here,subset accuracy is 0.125 and All feature accuracy is 0.98342.

```
Number of selected features: 1
[False False  True False False False False False False]
Subset accuracy: 0.12544891418858173
All Features Accuracy: 0.9834294892919607
```

Fig 9:Feature Selection using PSO methods.

The below figure 10 shows the Machine learning Evaluation metrics using Linear Regression.

```
from sklearn import metrics
print('Mean Absolute Error:', metrics.mean_absolute_error(y_test, y_pred))
print('Mean Squared Error:', metrics.mean_squared_error(y_test, y_pred))
print('Root Mean Squared Error:', np.sqrt(metrics.mean_squared_error(y_test, y_pred)))
print("Co-efficient of determination (R2 Score): ", metrics.r2_score(y_test, y_pred))
```

```
18.134791735073918
[ 0.13185982 -0.53577893 -0.00486174  0.67294637 -0.36198807  0.43374501
 -0.40910622 -0.09531441 -0.03697476  1.29936376]
Mean Absolute Error: 74.63067215537262
Mean Squared Error: 18480.570206778117
Root Mean Squared Error: 135.94326098331655
Co-efficient of determination (R2 Score):  0.9834294892919607
```

Fig 10: MachineLearning Performance using LinearRegression

The below fig 11 give details explanation of Model Prediction Perforance using statistical methods

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4s
DOI: https://doi.org/10.17762/ijritcc.v11i4s.6536
***Article Received**: 19 January 2023 **Revised**: 22 February 2023 **Accepted**: 01 March 2023*

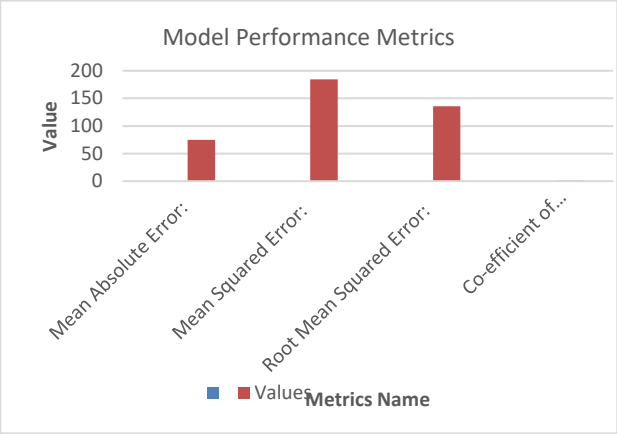

Fig 11: Model Performance comparision using statistical methods.

The below fig 12 shows descriptive analytics of crime Against Women Data-set.

```
[24] df.describe()
```

	2017	2018	2019	2021
count	25.000000	22.000000	28.000000	24.000000
mean	12525.120000	13073.136364	13974.428571	14225.708333
std	13230.682174	15195.818047	15132.845037	14849.609731
min	163.000000	75.000000	43.000000	130.000000
25%	1944.000000	836.500000	1494.500000	2973.000000
50%	8133.000000	7835.000000	8779.500000	8305.500000
75%	17909.000000	19435.500000	19736.000000	18678.750000
max	56011.000000	59445.000000	59853.000000	56083.000000

Fig 12:Descriptive analytics of CAW dataset

TABLE 1. Comparison of Model Performance

Data Set	Different Types of features Used	Types of Errors		
		MSE	***RMSE***	***R^2***
Crime against women in India	Extra Tree Classifier	0.321	0.28	0.95
	PSO	39.5961	42.562	0.923
	FIPSO	**0.256**	**0.2348**	**0.9889**

The above table 1 shows model performance of crime rate prediction.Here the proposed method FIPSO method return 98.89 % accuracy.This is highest value compare than others two methods.

Table 2. Comparison of Actual and Predicted Values with NCRB Reoprt

Data Set	NCRB and Predicted Values with			
	Row	***Column***	***NCRB Value***	***Predicted***
Crime against women in India	6	2	10986	10902
	71	4	20872	20181
	82	7	18547	18335

V. CONCLUSION

Crimes against women are a major problem in our country. The crime rate against women is constantly increasing. Crime prediction is key to solving this problem. Data analysis is a scientific method for analyzing and interpreting raw data results. Many data analysis methods are performed automatically, with algorithms processing information essential for human consumption. Predictive analytics is one of the well-known techniques of data analysis that helps predict future value. This paper proposed, LRSRI for missing value imputation and FIPSO method for future selection. Linear regression techniques implemented for crime rate prediction . The results of this algorithm are compared to the NCRB Crime Reporting Score. The proposed algorithms returned 98.89% accuracy with low mean squared error. In future,The results will help the Criminal Investigation Department to control the crimes against women in India.

REFERENCES

[1] Karmitsa, Napsu; Taheri, Sona; Bagirov, Adil; Makinen, Pauliina (2020). Missing Value Imputation via Clusterwise Linear Regression. IEEE Transactions on Knowledge and Data Engineering,(),1–1 doi:10.1109/TKDE.2020.3001694

[2] Ali, Azza & Abuelkheir, Mervat & Atwan, Ahmed & Elmogy, Mohammed. (2022). Missing Values Imputation Using Fuzzy K-Top Matching Value. Journal of King Saud University - Computer and Information Sciences. 35. 10.1016/j.jksuci.2022.12.011.

[3] Nadimi-Shahraki MH, Mohammadi S, Zamani H, Gandomi M, Gandomi AH. A Hybrid Imputation Method for Multi-Pattern Missing Data: A Case Study on Type II Diabetes Diagnosis. Electronics. 2021; 10(24):3167. https://doi.org/10.3390/electronics10243167

[4] Rahin Atiq, Farzana Fariha,Mutasim Mahmud and Sadman S. Yeamin,Kawser I. Rushee1 and Shamsur Rahim,"A Comparison of Missing Value Imputation Techniques on Coupon Acceptance Prediction",I.J. Information Technology and Computer Science, October 8, 2022 ,PP-15-25,DOI: 10.5815/ijitcs.2022.05.02

[5] Amelia Ritahani Ismail , Nadzurah Zainal Abidin , Mhd Khaled Maen, "Systematic Review on Missing Data Imputation Techniques with Machine Learning Algorithms for Healthcare", Journal of Robotics and Control,Volume 3, Issue 2, March 2022 ISSN: 2715-5072 DOI: 10.18196/jrc.v3i2.13133

[6] Muhammad Metwally Seliem "HandlingOutlierData as Missing Values by Imputation Methods: Application of Machine Learning Algorithms", *Turkish journal of Computer and mathematics Education",Vol 13,No 1,2022,PP- 273-286*

[7] Saba Bashir, Irfan Ullah Khattak, Aihab Khan, Farhan Hassan Khan, Abdullah Gani, Muhammad Shiraz, "A Novel Feature Selection Method for Classification of Medical Data Using Filters, Wrappers, and Embedded

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4s
DOI: https://doi.org/10.17762/ijritcc.v11i4s.6536
***Article Received**: 19 January 2023 **Revised**: 22 February 2023 **Accepted**: 01 March 2023*

Approaches", Complexity, vol. 2022, Article ID 8190814, 12 pages,2022. https://doi.org/10.1155/2022/8190814

[8] Kaushalya Dissanayake, Md Gapar Md Johar, "Comparative Study on Heart Disease Prediction Using Feature Selection Techniques on Classification Algorithms", Applied Computational Intelligence and Soft Computing, vol. 2021, Article ID 5581806, 17 pages, 2021. https://doi.org/10.1155/2021/5581806

[9] Tarneem Elemam, Mohamed Elshrkawey, "A Highly Discriminative Hybrid Feature Selection Algorithm for Cancer Diagnosis", The Scientific World Journal, vol. 2022, Article ID 1056490, 15 pages, 2022. https://doi.org/10.1155/2022/1056490

[10] NamJung Kim,KyongMin Min,"Optimal machine learning feature selection for assessing the mechanical properties of a zeolite framework",Physical Chemistry Chemical Physics,Issue 44, 2022