Risk Minimization Process on Crime Cluster Data

Dileep Kumar Kadali[1], R.N.V. Jagan Mohan[2], and M. Chandra Naik[3]

[1]Research Scholar, [2]Professor,[3]Professor

[1]Dept. of CSE, GIET University, Gunupur, Odisha, India-765022.

[2]Professor, Sagi Rama Krishnam Raju Engineering College, Bhimavaram, India-534 204.

[3]Professor, Dept. of CSE, GIET University, Gunupur, Odisha, India-765022.

Abstract - Over the past few years, machine learning system methods can play a major role in particularly applying risk minimization methods to an increasing design of systems. Crime plays a panic role in society because it is very risky in human life. It is reasoned to identify the risk of crime from existing, that the availability of learning methods has been a curtail factor in the recent success of crime rate applications. This paper proposed crime risk minimization techniques using machine learning methods based on inductive principles. Typically, in machine learning models, a summed-up model should be chosen from limited group datasets with the subsequent issue of overfitting the singular's way of behaving is perceived and broken down with statistician science since exercises or association are taking a converging individual security and rundown of their clients is the model turning out to be too firmly customized disposition of the preparation set and sum up simply to new information. The standard location of this issue is by adjusting a model's intricacy against its prosperity at fitting the preparation information.

Index Terms - Machine Learning, Structural Risk Minimization, Cluster, Security, etc.

I. Introduction

A cluster is a collection of things that are similar to the same class. In further words, the correlated things are clustered in one cluster, and dissimilar things are collected in the alternative cluster. Clustering is the technique of producing a set of abstract objects into classes of parallel things. A cluster of data things can be frozen as one group although in doing cluster analysis; Initially partition the set of data into clusters based on data similarity and then allot the labels to the clusters. The focal benefit of clustering over classifications is that it is flexible to changes and helps single out useful features that distinguish different groups. Clustering analysis is approximately used in many applications such as outline data analysis, and machine learning. It can also support discovering distinct clusters in their crime base. They can describe their crime person's clusters based on the patterns. In the field of criminology, facts can be used to arise criminal data with similar functionalities and gain insight into structures essential to people. Clustering also supports the credentials of zones of similar land usage in a ground reflection database. It also supports the credentials of clusters of patterns in a city. It can also help in classifying the data on evidence detection. It is also used in outlier detection applications such as the detection of deception. By way of a Data mining function, cluster analysis serves as a tool to gain insight into the distribution of data to observe the appearances of each cluster. Most data mining techniques would cluster crime data differently than statistical. If applied K-Means to the same data Modeler would provide a different result from the Statistics' K-means tool. If have access to Modeler and may find it enlightening to use the 'auto clustering' feature. The cluster tool will run multiple algorithms against the data. We can then look at the fit of each algorithm. It can use good analytic practices then we can use a model subset of the data, and test the goodness of fit. To run the different data sets in the same manner. It is a good way to get a feel for how the different algorithms work with different datasets. Annotation is the difference between supervised and unsupervised machine learning. It is to understand to mathematics is important, as understanding how to model data and how to run an analysis.

II. Absolute Error Criterion

The squared error criterion has one significant drawback. It is heavily affected by the presence of (data points with extreme values) in the dataset. The distance of an outlier point from its cluster will be quite large. The square of this distance will be even larger. To avoid the above drawback, squaring the distances of points from their cluster centers can be avoided. The resulting measure is known as the absolute error criterion. It is expressed as

$$E = \sum_{i=1}^{N} \sum_{p \epsilon C} d(p, \overline{p_i})$$

Where d is the distance function and p_i is the center of the cluster C_i. This measure of the clustering quality is relatively immune to the presence of outliers.

III. Structural Risk Minimization

Structural Risk Minimization (SRM) is a learning model to fit a moral simplification concert over an input cluster data set. The input data set is considered as. It is to produce a label fit model of capacity with tradeoff space complexity for training input clusters. The quality of the fitting model will be generated as

a. Overwhelming a priori familiarities of the field, select a class function with multinomials of step N, neural networks taking N hidden layer neurons, a customary of keys with N nodes, or fuzzy logic models overriding N rules.
b. Split the class of function having a sub-class functions hierarchy of draw close subgroups in the directive of increasing complexity.
c. Perform observed risk minimization on each subset.
d. Select a model in the succession of empirical risk and VC self-confidence is nominal.

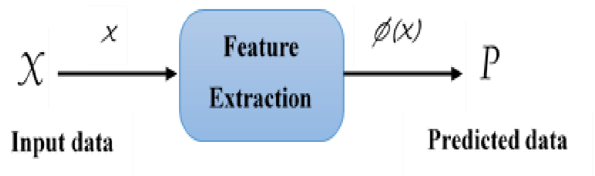

Fig. 1. Feature Extraction for Input Data and Predicted Data

The structural risk minimization model is a rule that is at least partially 'used' in all machine-learning approaches, meanwhile overfitting is often to be occupied into class: reducing the complexity of the model is a moral mode to edge overfitting.

- To clearly have a limit for the complexity and it is essential because increasing the complexity is one of the parts of the classification learning procedure.
- Take an informal stick of their complexity and are part of the study to the related procedure.
- Without this principle, the parsing implication would stand for both stupid and perfect syntax is a list of all likelihood words, so all non-trivial algorithms having minimum acknowledges this principle parsing.
- Decision trees form their particular notion of entropy.
- Clusters can be simply counted or kind of 'use' the principle intrinsically or have a fixed number of clusters and in that case, you apply the principle at a higher level.

IV. Risk Minimization Process

The main objective of knowledge is typically to invent a new model which carries a decent simplified presentation and concluded a basic circulation of the data. Study a contribution input space is X and output space is P.
Take up the pairs (X × P) $\epsilon\ X \times P$ are random variable quantities whose combined distribution is D_{XP}. It is main aim to find a predictor f: X →Y, which reduces the probably expected risk:

$$D\ \{f(X) \neq P\} = E_{(X,P)} \sim D_{XP}\ [\delta\ \{f(X) \neq P\}]$$

We get here δ(z) = 1 when z is true, otherwise δ(z) =0.

In training, initially, we have n pairs of training examples (X_i, P_i) tired identically and autonomously from D_{XP}. Founded on these models, the empirical risk can be well-defined as

$$\frac{1}{n}\sum_{i=1}^{n} \delta(f(X) \neq P)$$

Picking a particular function f by minimalizing an empirical risk frequently leads to overfitting. To improve this procedure, the knowledge of structural risk minimization (SRM) is to service an unlimited classification of models F_1, F_2 .... F_n. through growing measurements. Now respectively F_i is an established of functions, e.g., multinomials with degree 3. We minimalize the empirical risk in each model by means of a penalty aimed at the measurements of the model:

$$F_n = \underset{f\epsilon F_i\ i\epsilon n}{argmin} \frac{1}{n}\sum_{i=1}^{n} \delta(f(X) \neq P) + capacity(F_i, n)$$

Where measurements (F_i, n) identify the complication of model F_i in the setting of the specified training set. For instance, it equals 2 after F_i is the set of multinomials with degree 2. In further words, at what time trying to decrease the risk on the training dataset, we desire a predictor after a humble model. Memo the consequence is slow on the model F_i, non the predictor *f*. This is dissimilar beginning the regulation context, e.g., support vector machines, which punishes the difficulty of the classifier.

V. Experimental Result

For the structural risk minimization model was applied purpose study an input-space X and output-space P to four criminal data sets are Murder, Rape, Robbery and Auto Theft, rule that is used in machine-learning approaches, meanwhile overfitting is often to be occupied into class: reducing the complexity of the model is a moral mode to edge overfitting.

TABLE-I

Units for Crime Murder Actual & Prediction

SNo	City Names	Murder	Prediction
1	Bhimavaram	16.5	0.9999999
2	Narasapuram	4.2	0.985226
3	Tadepalligudem	11.6	0.9999908
4	Eluru	18.1	1
5	Vijayawad	6.9	0.9989932
6	Kakinada	13	0.9999977
7	Tuni	2.5	0.9241418
8	Ravulapalem	3.6	0.973403
9	Rajahmundy	16.8	0.9999999
10	Amalapuram	10.8	0.9999796
11	Palakollu	9.7	0.9999387
12	Visakhapatnma	10.3	0.9999664
13	Razole	9.4	0.9999173
14	Gudivada	5	0.9933071
15	Gunturu	5.1	0.9939402
16	Annavaram	12.5	0.9999963

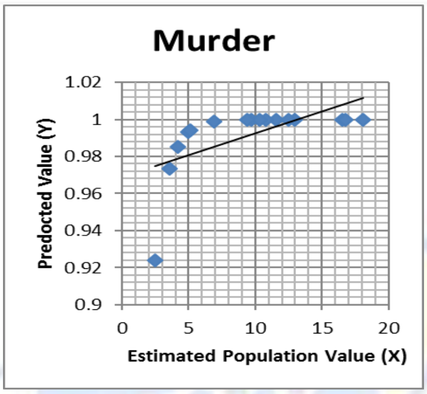

Fig. 2. Fitting for Crime Murder Estimated data and Predicted data

TABLE-II

Units for Crime Rape Actual & Prediction

SNo	City Names	Rape	Prediction
1	Bhimavaram	24.8	1
2	Narasapuram	13.3	0.9999983
3	Tadepalligudem	24.7	1
4	Eluru	34.2	1
5	Vijayawad	41.5	1
6	Kakinada	35.7	1
7	Tuni	8.8	0.9998493
8	Ravulapalem	12.7	0.9999969
9	Rajahmundy	26.6	1
10	Amalapuram	43.2	1
11	Palakollu	51.8	1
12	Visakhapatnma	39.7	1
13	Razole	19.4	1
14	Gudivada	23	1
15	Gunturu	22.9	1
16	Annavaram	27.6	1

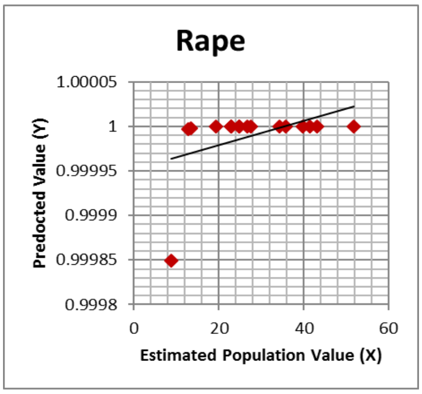

Fig. 3. Fitting for Crime Rape Estimated data and Predicted data

TABLE-III

Units for Crime Robbery Actual & Prediction

SNo	City Names	Robbery	Prediction
1	Bhimavaram	106	1
2	Narasapuram	122	1
3	Tadepalligudem	340	1
4	Eluru	184	1
5	Vijayawad	173	1
6	Kakinada	477	1
7	Tuni	68	1
8	Ravulapalem	42	1
9	Rajahmundy	289	1
10	Amalapuram	255	1
11	Palakollu	286	1
12	Visakhapatnma	266	1
13	Razole	522	1
14	Gudivada	157	1
15	Gunturu	85	1
16	Annavaram	524	1

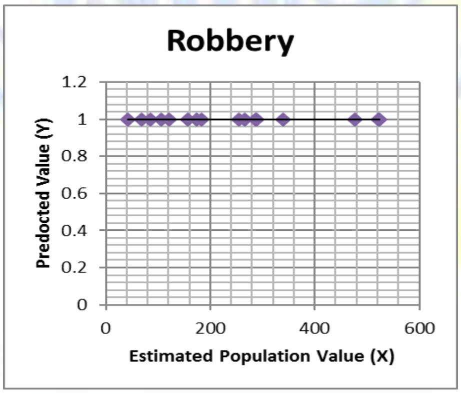

Fig. 4. Fitting for Crime Robbery Estimated data and Predicted data

TABLE-IV

Units for Crime Auto Theft Actual & Prediction

SNo	City Names	Auto Theft	Prediction
1	Bhimavaram	494	1
2	Narasapuram	954	1
3	Tadepalligudem	645	1
4	Eluru	602	1
5	Vijayawad	780	1
6	Kakinada	788	1
7	Tuni	468	1
8	Ravulapalem	637	1
9	Rajahmundy	697	1
10	Amalapuram	765	1
11	Palakollu	862	1
12	Visakhapatnma	776	1
13	Razole	848	1
14	Gudivada	488	1
15	Gunturu	483	1
16	Annavaram	793	1

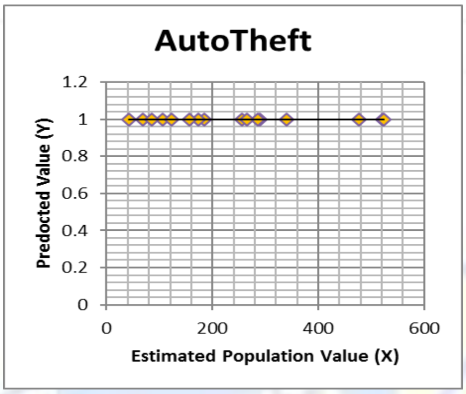

Fig. 5. Fitting for Crime Auto Theft Estimated data and Predicted data

VI. CONCLUSION

Our proposed work is on the classification of crime data clusters with the help of structural risk minimization in this regard estimated the error rate using absolute error on crime data clusters. The experimental result is measured on crimes data clusters are Murder, Rape, Robbery and Auto Theft comprises Finally we conclude that minimize the error rate on crimes data.

VII. REFERENCES

[1] RNV Jagan Mohan, Dileep Kumar Kadali and et al. "Crime data optimization using Neutrosophic logic based game theory", Concurrency and Computation: Practice and Experience, 2022.

[2] Kadali, D.K., Chandra Naik, M., Jagan Mohan, R.N.V. (2023). “Estimation of Data Parameters Using Cluster Optimization” In: Goswami, S., Barara, I.S., Goje, A., Mohan, C., Bruckstein, A.M. (eds) Data Management, Analytics and Innovation. ICDMAI 2022. Lecture Notes on Data Engineering and Communications Technologies, vol 137. Springer, Singapore. https://doi.org/10.1007/978-981-19-2600-6_23.

[3] Dileep Kumar Kadali, and R.N.V.Jagan Mohan “Shortest Route Analysis for High level Slotting Using Peer-to-Peer” in Taylor and Francis Group, The Role of IoT and Blockchain: Techniques and Applications CRC press,2022.

[4] Dileep Kumar Kadali, R.N.V.Jagan Mohan and M. Srinivasa Rao “Cluster Optimization for Similarity Process Using De-Duplication” in IJSRD -International Journal for Scientific Research & Development| VOL. 4, Issue 06, AUG-2016 | ISSN (online): 2321-0613.

[5] Dileep Kumar Kadali and R.N.V.Jagan Mohan “Optimizing the Duplication of Cluster Data for Similarity Process” in ANU Journal of Physical Science| VOL-2, JUN-DEC-2014 | ISSN: 0976- 0954.

[6] Dileep Kumar Kadali, R.N.V.Jagan Mohan and Y. Vamsidhar “Similarity based Query Optimization on Map Reduce using Euler Angle Oriented Approach” in International Journal of Scientific & Engineering Research, Volume 3, Issue 8, August-2012, ISSN 2229-5518.

[7] Jagan Mohan R.N.V. and Subbarao. R. and Raja Sekhara Rao K., “Efficient K-Means Cluster Reliability on Ternary Face Recognition using Angle Oriented Approach”, Published In the Proceedings of International Conference on Advances in Communication, Navigation and Signal Processing Technically Co-Sponsored by IEEE, Hyderabad Section, March 17th-18th, 2012, Dept of ECE, Andhra University College of Engineering (A).

[8] Dileep Kumar Kadali, R.N.V. Jagan Mohan, M. Chandra Naik, Unsupervised based crimes cluster data using decision tree classification. J. Solid State Technol. 63(5), 5387–5394, 2020.

[9] W.K.Li, W.L.Wang, and L.Li “Optimization of water resources utilization by multi-objective moth-flame algorithm” in Water Resources Management, vol. 47, no:10, pp. 3303–3316, 2018.

[10] Dileep Kumar Kadali, R.N.V. Jagan Mohan, M. Chandra Naik, "Crime Cluster Reliability Through Game Theory and Neutrosophic Logic Using Three-Stage Model" in Journal of Data Acquisition and Processing, 2023, 38 (1): 2689-2697 .

[11] X. Bi and C. Wang "A niche-elimination operation-based NSGA-III algorithm for many-objective optimization" in Applied Intelligence, vol. 48, no:1, pp. 118–141, 2018.

[12] S.Mahdavi, S.Rahnamayan, and K.Deb "Opposition based learning: a literature review," Swarm and evolutionary computation, vol. 39, pp. 1–23, 2018.

[13] S.Mirjalili, P.Jangir, and S.Saremi "Multi-objective ant lion optimizer: a multi-objective optimization algorithm for solving engineering problems" in Applied Intelligence, vol. 46, No:1, pp. 79–95, 2017.

[14] Y. Xiang, Y. Zhou, M. Li, and Z. Chen "A vector angle-based evolutionary algorithm for unconstrained many-objective optimization" in IEEE Transactions on Evolutionary Computation, vol. 21, no. 1, pp. 131–152, 2017.

[15] S. Jiang and S. Yang "A strength pareto evolutionary algorithm based on reference direction for multiobjective and many-objective optimization" in IEEE Transactions on Evolutionary Computation, vol. 21, no. 3, pp. 329–346, 2017.

[16] Dileep Kumar Kadali, R.N.V. Jagan Mohan, M. Chandra Naik, "Empirical Analysis on Uncertain Crime Data using Hybrid Approaches", CIMS, vol. 28, no. 12, pp. 1223–1237, Dec. 2022.

[17] W.Wang, S.Ying, L.Li, Z.Wang, and W. Li "An improved decomposition-based multiobjective evolutionary algorithm with a better balance of convergence and diversity" in Applied Soft Computing, vol. 57, pp. 627–641, 2017.

[18] C.Zhu, L.Xu, and E.D.Goodman, "Generalization of pareto-optimality for many-objective evolutionary optimization" in IEEE Transactions on Evolutionary Computation, vol. 20, no. 2, pp. 299–315, 2016.

[19] Y.Yuan, H.Xu, B.Wang, and X.Yao, "A new dominance relation-based evolutionary algorithm for many-objective optimization" in IEEE Transactions on Evolutionary Computation, vol. 20, No:1, pp. 16–37, 2016.

[20] B.Li, K.Tang, J.Li, and X.Yao "Stochastic ranking algorithm for many-objective optimization based on multiple indicators" in IEEE Transactions on Evolutionary Computation, vol. 20, No:6, pp. 924–938, 2016.

[21] Y.Yuan, H.Xu, B.Wang, B.Zhang and X.Yao "Balancing convergence and diversity in decomposition-based many-objective optimizers" in IEEE Transactions on Evolutionary Computation, vol. 20, no. 2, pp. 180–198, 2016.

[22] R.Cheng, Y.Jin, M.Olhofer and B.Sendhoff "A reference vector guided evolutionary algorithm for many-objective optimization" in IEEE Transactions on Evolutionary Computation, vol. 20, No:5, pp. 773–791, 2016.