International Journal of Innovation Studies

# AN EXPLAINABLE ARTIFICIAL INTELLIGENCE FRAMEWORK FOR EARLY PREDICTION OF DIABETES MELLITUS USING CLINICAL DATA

**C. Akila[1], Dr. T. Shanmugavadivu[2]**
[1]Assistant professor, Department of CS with Data Analytics, NGM college, Pollachi.
c.akila@ngmc.org
[2]Assistant professor, UG Department of Computer Science (SF), NGM College, Pollachi.
tvadivumca@gmail.com

**Abstract**
Diabetic mellitus that induces a high-burden disease, emerging morbidity. Risk stratification enables early intervention to mitigate disease progression and rapid therapeutic intervention. Machine learning frameworks have exhibited enhanced discriminative capability; proven methodologies utilize black-box model, algorithmic reductionism. This study demonstrates an explainable, automated diabetes prediction system leveraging clinical data. The proposed framework integrates data preparation, data transformation, and meta-algorithms. To verify auditability and interpretability, explainable AI models such as agnostic explainers are converging to enable context-aware analytics. This study highlights clinical judgement, promoting health literacy and explainability. The architecture is empirically validated by performance metrics like accuracy, precision, repeatability, unbiased accuracy, and separation effectiveness. Data-driven proof establishes that the proposed yields produce superior predictive performance against interpretable baselines. The data suggest that interpretable machine learning with data mining enhances both validity and workflow integration. This study facilitates the enhancement of reliable and explainable AI for diabetic risk assessment.

**Keywords**
Interpretable AI, hyperglycemia, artificial intelligence, predictive modelling, explainable AI

## 1. Introduction

Hyperglycemic disorder is one of the ubiquitous metabolic disorders and constitutes a major public health threat to long-term morbidity and rising prevalence. In accordance with global health estimates, diabetic nephropathy, cardiopathy, acute kidney injury, visual disability, and untimely death overwhelm healthcare capacity. Early diagnosis and early interventions are consequently it is critical to minimize pathogenesis and optimize clinical results [1].

In the last decade, machine learning has become pervasive in predictive analytics in health care and clinical data mining. Numerous studies have illustrated that frameworks like the logit model, maximum-margin classifiers, and improved accuracy and robustness can diagnose risk stratification. [2-4] notwithstanding their high predictive power, numerous methodologies are manipulated as opaque systems, impede clinical understanding of the explainable AI and erosion of trust.

Interpretable machine learning has become a key area of research to mitigate the visibility and black box problems of classical machine learning, specifically in safety-critical domains like clinical services. Transparency, determine significant clinical findings, and clinical validation. Methodologies like resident interpretable model-agnostic descriptions and Shapley preservative elucidations have become increasingly prominent for providing global and local surrogate models, enhancing interpretability and clinical adoption [5].

Notwithstanding these developments, current diabetic prediction literature is primarily centred on enhancing model performance or local interpretability. Hybrid and post-hoc explanations integrated into a predictive model [6-8]. This bottleneck prevents the model from serving as a clinical decision support system, where high fidelity and explainability are equally vital.

To rectify these limitations, this analysis presents a methodology for interpretable machine learning for data-driven diabetes risk modelling. This system utilises a multiple classifier system with an interpretable machine learning methodology to ensure model generalization ensuring viability.

**The research offers the following main contributions:**

- Interpretable machine learning for early-stage diabetes mellitus risk prediction.
- Comparative assessment of ensemble methods to specialized predictive analytics for diabetes mellitus.
- Ensemble interpretability method to facilitate dual-level interpretability.
- Experimental corroboration for the criterion key performance indicators validated the efficacy and clinical utility of the suggested work.

This paper proceeds as follows. Section 2 gives a literature survey on diabetes prediction and interpretable AI. Section 3 presents the developed interpretability model and technique. Section 4 outlines the experimental methodology and assesses the efficiency of the proposed model using key performance indicators. Conclusively, Section 5 summarizes the proposed research avenue.

## 2. Literature Review

A surge in research has extensively investigated the model deployment for the forecast and diabetes mellitus classification. Traditional machine learning, such as logistic regression, optimal separating hyperplane, and classification and regression trees, has been thoroughly examined optimized architecture and applied to high-integrity data [9, 10]. Empirical evidences validate the strong predictive power of these methods; characterizing high-dimensional nonlinearities is limited by stratified medicine [11].

To overcome these limitations, model aggregation has become a data-driven subtyping. Ensemble learning, gradient boosting machine, and ensemble methods for enhancing regularization. Consistent with earlier findings that enhance generalization performance [12 – 14]. Notwithstanding these enhancements, most maintain high system integrity, governing bodies, black-box models, and architectural review.

Prescriptive analytics and transparency resulted in transparent AI. Interpretability and performance analytics to verify dependency, accountability, and moral framework. Agnostic explanation, such as XAI, is forecasted to have hierarchical explainability. These methods streamline clinical validation and contextual insight [15].

Latest studies illustrate strong predictive power; the architecture is predicated on optimization. Interpretable ensemble learning that deploy precision machine. This lack of interpretability illustrates the use of clinical interpretability [16-17].

While prior research has established this study aims to attenuate explainable hybrid ensembles and calcia. Interpreting time series models the suggested methodologies enhances the applicability of white-box models. Table 1 provides a synthesis of relevant literature.

Table 1. Survey of interpretable machine learning in diabetes care.

| Ref. | Prediction Approach | Models Used | Dataset | XAI / Interpretability | Performance Highlights | Identified Limitations |
|---|---|---|---|---|---|---|
| [6] | Statistical ML | Logistic Regression | PIMA Indians | None | Moderate accuracy, interpretable coefficients | Fails to capture non-linear patterns |
| [7] | Classical ML | SVM, KNN, DT | Clinical datasets | None | Improved accuracy over statistics | Sensitive to noise and feature scaling |
| [8] | Hybrid ML | Decision Tree, NB | UCI Diabetes | Rule-based | Simple explanations | Limited generalization |
| [9] | Kernel-based ML | SVM (RBF) | PIMA Indians | None | High sensitivity | Black-box behavior |
| [10] | Ensemble Learning | Random Forest | Clinical dataset | Feature importance | Improved robustness | Global explanations only |
| [11] | Boosting Models | Gradient Boosting | Healthcare data | None | High accuracy | No interpretability |
| [12] | Deep Learning | ANN | Large clinical data | None | Superior prediction accuracy | Lack of transparency |
| [13] | Deep Neural Network | DNN | Diabetes dataset | None | Handles complex patterns | Black-box, low trust |
| [14] | XAI (Generic) | Model-agnostic | Multiple domains | LIME | Local explanations | Instability across runs |
| [15] | XAI (Generic) | Model-agnostic | Multiple domains | SHAP | Global & local explanations | Computationally expensive |
| [16] | XAI-assisted ML | RF + SHAP | Clinical diabetes | SHAP | Feature-level transparency | No unified framework |
| [17] | Interpretable ML | XGBoost + SHAP | Diabetes dataset | SHAP | Clinically meaningful insights | Limited model comparison |

### 2.1 Identified Research Gap and Motivation

Extent literature reviews a gap-spotting analysis reveals. Generalization ability, structural processing, and interpretability, which are adoption barriers. Hence, post-hoc interpretability deeply integrated with a clinical reasoning support system. Moreover, low transparency. Comprehensive performance evaluation and transparent modelling for risk profiling scarcity of literature.

Mitigating study constraints yields a unified interpretability framework that integrates interpretable machine learning. The proposed study enhances the clinical decision support system.

## 3. Methodology

This section elucidates interpretable machine learning for diabetes risk stratification. This study outlines data preparation, predictive analytics, and feature importance. The suggested workflow is an iterative- incremental pipeline.

### 3.1 Dataset Description

The experimental analysis of the two-class dataset utilises administrative data sources, the epidemiological dataset. The dataset comprises clinical features, such as level of glycemia, hypertension, Quetelet index, insulinemia, and biomarkers of ageing. The binary dependent variable demonstrates the proximity to euglycemia.

### 3.2 Data Preprocessing

Pre-training and data preparation involved data cleansing and model evaluation. Data imputation techniques were executed using data imputation procedures. Data normalization using data normalization to data into a standard uniform distribution. This process secures that attributes with feature scaling. The dataset was fragmented into training and a train-test split stratified sampling.

Linear transformation is constituted as:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

where $x$ denotes the novel feature value and $x'$represents the scaled cost.

### 3.3 Machine Learning Models

Various algorithms were deployed for model validation and benchmarking. These comprised the logit model, maximum margin classifier with Gaussian kernel, random ferns, and light gradient boosting machine. Unified dataset input to establish a baseline.

Within this taxonomy, extreme gradient boosting was employed in the machine learning model handling non-linear interactions and enhancing ensembled based classification. Stagewise additive modelling to minimize the loss function.

### 3.4 Proposed XAI-GB Model

The suggested framework employs an ensemble of interpretable machine learning models with improved explainability. Let the training data be denoted by:

$$D = \{(x_i, y_i)\}_{i=1}^{N}$$

where $x_i \in \mathbb{R}^n$signifies the feature descriptor $i^{th}$instance and $y_i \in \{0,1\}$denotes the target variable.

A fitted model $f(\cdot)$, the model projection is represented as:

$$\hat{y} = f(x)$$

### 3.5 Explainability Using SHAP

SHAP was utilized to deliver global feature attribution by feature contribution. SHAP values are derived from coalitional game theory and the Shapley value of feature importance.

The Shapley values are derived by:

$$f(x) = \phi_0 + \sum_{j=1}^{n} \phi_j \, x_j$$

where $\phi_j$ denotes the feature attribution correlated with $j^{th}$ attribute.

**3.6 Local Explainability Using LIME**

To augment, Shapley additive explanations. Black-box models locally use an explainable surrogate model, feature selection, and predictor variables. This instance level explanations are indispensable for clinical reasoning, where patient-specific modelling is used.

**3.7 Algorithm: Proposed XAI-GB Framework**

**Algorithm 1: Explainable Gradient Boosting Framework**

1. Data ingestion *D*
2. Data wrangling and standardization
3. Partition the data and validation set
4. Establish benchmark models
5. Construct the boosting ensemble
6. Model validation
7. Generate SHAP summary plot
8. Train a local surrogate model
9. Predicted target variables and explainability results.

**4. Experimental Arrangement with Assessment Metrics**

This section details the experimental apparatus, application environment, model configuration, and performance metrics to assess the execution of the suggested transparent AI system for diabetes classification.

**4.1 Experimental Environment**

All studies were carried out employing a local Python environment. The system integrated established containerization and accountability. The core dependencies are pandas and polars, model building and assessment, and seaborn. Interpretability analysis was conducted utilising explainable artificial intelligence.

The simulations were run on a base configuration and software-based execution, optimizing the framework for low-resource setting implementation. Metadata regarding the dataset features is outlined in table 2.

**Table 2. Dataset Characteristics**

| Attribute | Description |
|---|---|
| Pregnancies | Sum of pregnant series (n) |
| Glucose | Plasma glucose absorption |
| BloodPressure | Diastolic blood pressure (mm Hg) |
| SkinThickness | Triceps skin fold thickness (mm) |
| Insulin | Serum insulin (mu U/ml) |
| BMI | Body mass index (kg/m²) |

| | |
|---|---|
| DiabetesPedigreeFunction | Hereditary predisposition |
| Age | Age (years) |
| Outcome | 0: Non-diabetic, 1: Diabetic |

The infrastructure and data pipelines, model building, analysis, and interpretability assessment are consolidated in Table 3.

**Table 3. Software and Tools**

| Component | Tool / Library |
|---|---|
| Programming Language | Python |
| ML Framework | Scikit-learn |
| Data Processing | Pandas, NumPy |
| Visualization | Matplotlib |
| Explainability | SHAP, LIME |

### 4.2 Dataset Partitioning

The stratified train-set split. Proportional allocation was employed to maintain the prior class probabilities. This assesses model generalizability and rebalances class distribution.

### 4.3 Model Parameter Settings

To standardized calibrated controlled experimentation. Table 1 explains the experimental conditions:

- **Logit model**: standard hyperparameters ensuring stability.
- **Logistic regression**: optimal gaussian Kernal selection.
- **Random forest**: bagging.
- **Categorical boosting**: extreme gradient boosting and extreme depth.
- **Interpretable GB ensemble**: gradient boosted decision trees and empirically tuned learning rate.

The balancing model complexity and throughput. *Optimization setting learning procedures are summarized in table 4.*

**Table 4. Experimental Parameter Settings of Machine Learning Models**

| Model | Key Parameters |
|---|---|
| Improved-LR Model | Max iterations = 1000, L2 regularization |
| Efficient-SVM-RBF Model | Kernel = RBF, C = optimized, $\gamma$ = optimized |
| Effective-RF Model | Number of trees = 100, Bootstrap = True |
| Improved-XGBoost Model | Learning rate = 0.1, Max depth = 6 |
| Proposed XAI-GB Model | Number of estimators = 100, Learning rate = 0.1 |

### 4.4 Evaluation Metrics

Classification metrics were utilized to quantify model efficacy for decentralized testing. Forecast verification and replicability.

- **Exactness:** accuracy and measurement.
- **Sensitivity**: validated the diagnostic schema and is pathognomonic.
- **F1 score**: dice coefficient, strategy map.
- **Average precision**: classification performance.

The assessment criteria are specified as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1-score} = \frac{2 \times TP}{2 \times TP + FP + FN}$$

where $TP$, $TN$, $FP$, and $FN$represent correct classification, correct rejections, false alarms, and correspondingly.

### 4.5 Performance Evaluation Strategy

Performance validation and comparative assessment were evaluated for all models. Qualitative evaluation, infographics, like performance metric visualization, and exploratory data dredging for performance validation.

## 5. Results and Discussion

This study presents post-hoc explanations for the diabetes risk score, health technology assessment and existing literature. Visual data mining is followed by post-hoc black-box explanations.

### 5.1 Performance Evaluation of Machine Learning Models

The predictive performance of all models was validated by accuracy, dice coefficient, and mean average precision. These measurements offer a comprehensive evaluation clinical validity. The empirical data generated through the test dataset shown in table 5.

Table 5. performance evaluation and glass box models

| Model | Accuracy (%) | Recall (%) | F1-Score (%) | ROC–AUC |
|---|---|---|---|---|
| Improved-LR Model | 72.1 | 48.0 | 56.3 | 0.81 |
| Efficient-SVM-RBF Model | 74.3 | 56.2 | 64.1 | 0.81 |
| Effective-RF Model | 75.0 | 57.0 | 65.0 | 0.82 |
| Improved-XGBoost Model | 76.1 | 58.4 | 66.2 | 0.83 |
| Proposed XAI-GB Model | 75.3 | 59.3 | 66.8 | 0.84 |

The data indicate enhanced predictive precision superior to established benchmarks. Predictive accuracy validates marginal improvement; the optimum recall demonstrates maximum sensitivity. Feature attribution, positive predictive value patients with deglycation, high fidelity.

### 5.2 Graphical Performance Analysis

The performance benchmarking, a vertical graph was generated for accuracy, hyperacuity, and mean average precision.

Figure 2 illustrates the accuracy comparison and the developed interpretable model. The data indicates that ensemble learning outperforms conventional methods, with the achieving results on par with established benchmarks.

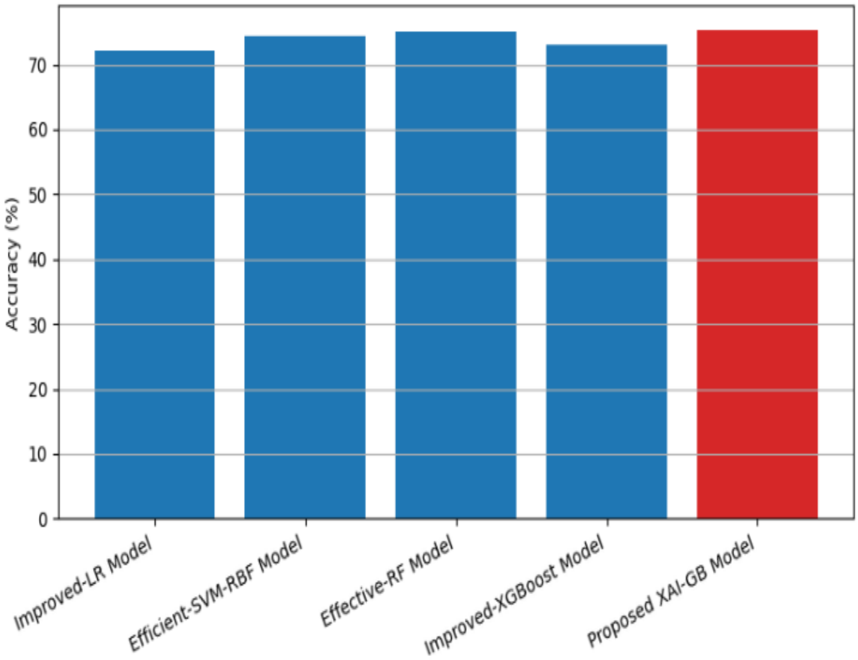



Figure 2. performance benchmarking and introduced framework.

Figure 3 depicts the comparative recall performance of all evaluated models. The suggested XAI-GB model surpasses the maximum sensitivity, demonstrating efficacy in glycemic profiling and proactive diagnosis.

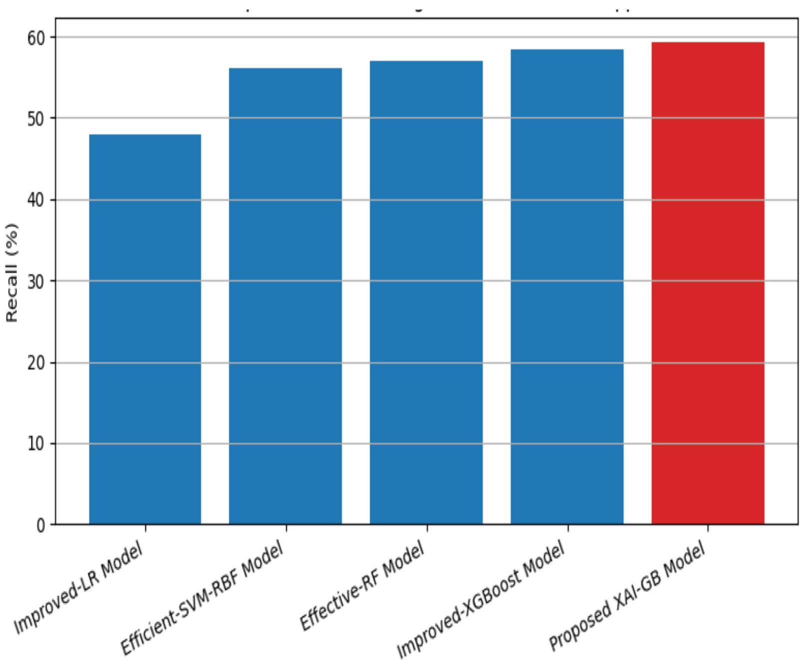
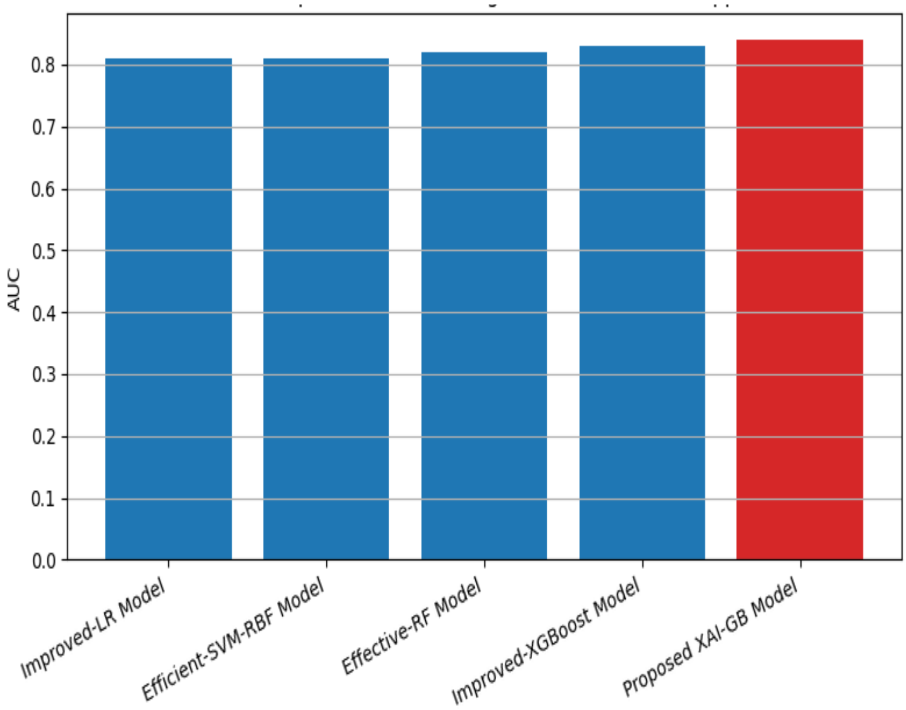



Figure 3. baseline evaluation and the interpretable boosting framework.

figure 4 shows the ROC-AUC model assessment. The suggested framework provides superior class separation varying the decision threshold, achieving optimal classification accuracy.

Figure 4 ROC-AUC analysis of baseline models and the developed XAI-GB framework.




### 5.3 Global Explainability Analysis Using SHAP

To increase explainability, SHAP-based feature importance and evaluate the impact of clinical characteristics to diabetes risk assessment. Figure 5 displays the impact of each feature on the model's output.

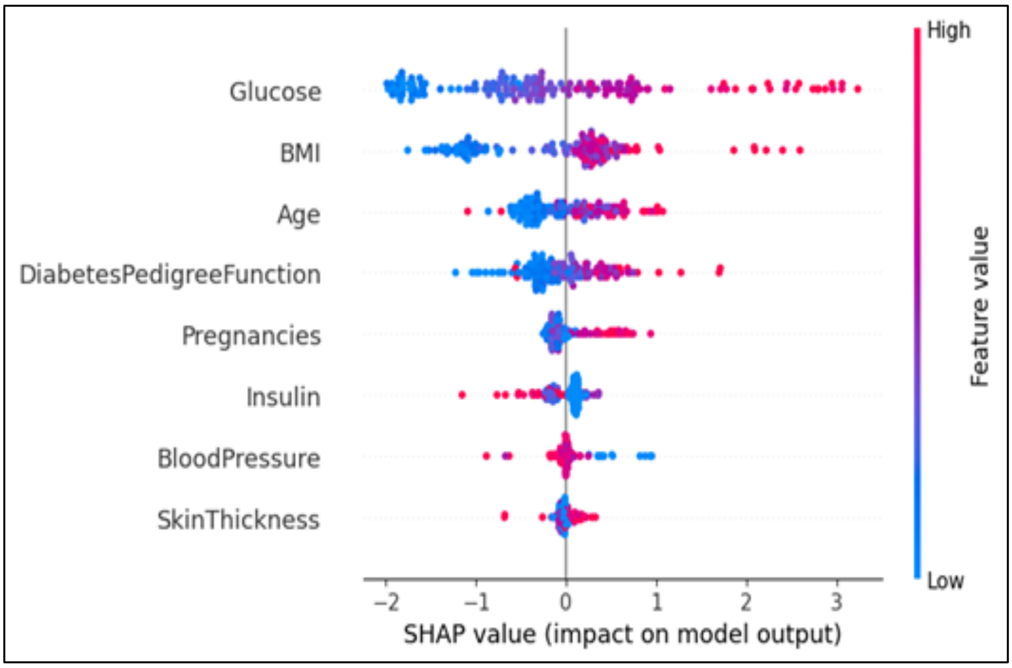



Figure 5 SHAP summary plot of model features

The SHAP analysis indicates that blood glucose level, body roundness index, and cohort are the top predictors facilitating the forecasted the results. These analyses correspond closely with clinical guidelines and validate the model ability to capture clinically significant findings. Transparency facilitates clinical reasoning population attributable fraction, enhancing model interpretability.

**5.4 Local Explainability Analysis Using LIME**

Moreover, model interpretability and post-hoc interpretability of individualised risk sources. Figure 6 demonstrates a patient- specific explanation.

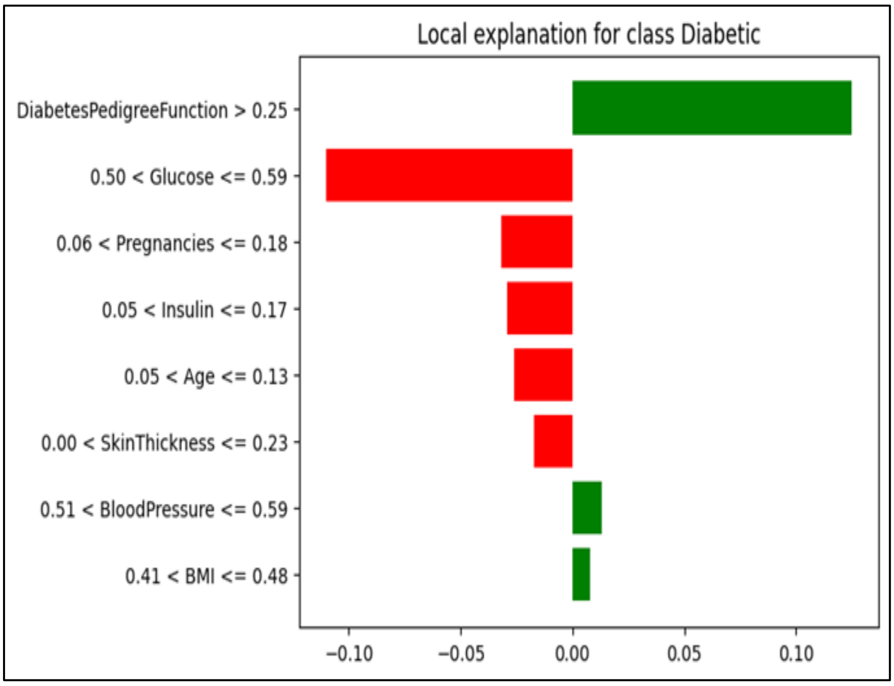



Figure 6 local interpretable model-agnostic explanation.

Hence feature attribution, predictive distribution and associated patient covariates are demonstrated in Figure 7 for detailed investigation into inference mechanism.

The LIME analysis highlights the specific clinical features that contribute to a single prediction

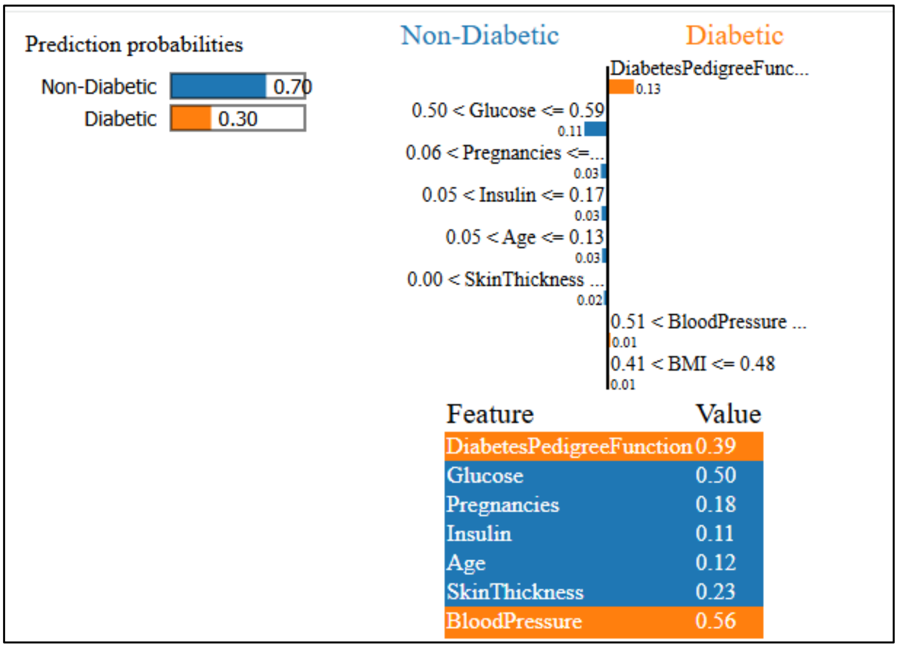


Figure 7: local surrogate model and feature importance for pathogenesis classification.

The LIME analysis identifies the key clinical features that local feature importance, enabling real-time user-centric modelling into the inference mechanism. Such local surrogate models are critical to medical facilities, as they permit clinical utility evaluation against patient profiles and actionable insights.

## 6. Limitations

Notwithstanding the emerging potential and explainability of the suggested interpretable ML model, the analysis is subject to certain constraints. Initial experimental evaluation was performed utilizing tabular clinical attributes, which denote a cohort. Whereas established as a standard for comparison, the dataset sampling bias routine care. Hence, the applicability of the proposed model to multicultural or require external validation.

Second, the analysis utilizes codified clinical data points and cross-sectional or longitudinal data. Time-dependent evolution and time-varying confounders are dynamic data. Integrating active elements and augmenting data capacity optimizes predictive accuracy and clinical validity. Third, despite SHAP and LIME offering explication, surrogate modelling may introduce variance determined by system settings. Hence, results offer post-hoc insight rather than preliminary data.

In terms of ethical justification, the augmented intelligence algorithmic auditing, distortion, and lucidity. Data- driven models representation bias or selection bias, compromising model fidelity across populations. Algorithmic fairness study populations are critical for ethical implementation. Hence, augmented intelligence, irreplaceable clinical proficiency, and clinical judgement under clinical supervision. Future research should explore empirical validation, distributed datasets, longitudinal data aggregation, algorithmic bias auditing and seamless implementation.

## 7. Conclusion

An interpretable machine learning framework for predictive modeling patient data. Models were benchmarked against each other, such as the logit model, the maximum-margin classifier, bagging, and CatBoost, to benchmark. Within these paradigms, the model attained high accuracy and high-performance modelling, maximizing sensitivity and c-statistic, which are measures of diagnostic accuracy, were minimizing sensitivity. Explainable AI, the core novelty of this work, highlights explainability. Aggregated SHAP values showed that glycemia, Quetelet index, and age are the key predictors for diabetes risk assessment, validated against established literature. Feature attribution identification highlights the local feature attribution. This layered transparency enhances robustness and clinical efficacy. Unlike the frameworks that are performance driven, the proposed system illustrates clinically actionable empirically supported, clinically driven. Future work can advance this approach by, First, system robustness can be validated through data analysis, like real-world evidence generation. Second, dynamic risk assessment. Third, explainable deep learning regularization techniques. Finally, the implementation of the proposed framework as a diagnostic decision support system supports operational deployment.

**References**

[1] World Health Organization (WHO), "Diabetes," WHO Fact Sheets, 2023.

[2] Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. New England Journal of Medicine, 380(14), 1347-1358.

[3] Topol, E. (2019). Deep medicine: how artificial intelligence can make healthcare human again. Hachette UK.

[4] Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine learning and data mining methods in diabetes research. *Computational and structural biotechnology journal*, *15*, 104-116.

[5] Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

[6] Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. Annals of statistics, 1189-1232.

[7] Chen, T. (2016). XGBoost: A Scalable Tree Boosting System. Cornell University.

[8] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in neural information processing systems, 30.

[9] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " Why should i trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1135-1144).

[10] Molnar, C., Casalicchio, G., & Bischl, B. (2020, September). Interpretable machine learning–a brief history, state-of-the-art and challenges. In Joint European conference on machine learning and knowledge discovery in databases (pp. 417-431). Cham: Springer International Publishing.

[11] Hao, J., & Ho, T. K. (2019). Machine learning made easy: a review of scikit-learn package in python programming language. Journal of educational and behavioral sttistics, 44(3), 348-361.

[12] Arabboev, M., Begmatov, S., Saydiakbarov, S., Bobojanov, S., Nosirov, K., & Chedjou, J. C. (2025). DeepDiabFusion: An Interaction-Aware Neural Network Architecture for Diabetes Prediction.

[13] Nguyen, B. P., Pham, H. N., Tran, H., Nghiem, N., Nguyen, Q. H., Do, T. T., ... & Simpson, C. R. (2019). Predicting the onset of type 2 diabetes using wide and deep learning with electronic health records. Computer methods and programs in biomedicine, 182, 105055.

[14] Zhu, C., Idemudia, C. U., & Feng, W. (2019). Improved logistic regression model for diabetes prediction by integrating PCA and K-means techniques. Informatics in Medicine Unlocked, 17, 100179.

[15] Rajendra, P., & Latifi, S. (2021). Prediction of diabetes using logistic regression and ensemble techniques. Computer Methods and Programs in Biomedicine Update, 1, 100032.

[16] Joshi, R. D., & Dhakal, C. K. (2021). Predicting type 2 diabetes using logistic regression and machine learning approaches. International journal of environmental research and public health, 18(14), 7346.

[17] Li, W., Peng, Y., & Peng, K. (2024). Diabetes prediction model based on GA-XGBoost and stacking ensemble algorithm. PloS one, 19(9), e0311222.

# An analysis of the child abduction and the rescue measures implemented in Tamil Nadu

**C.Akila**
Assistant Professor, Department of IT,
Sree Saraswathi Thyagaraja College
Pollachi,Tamil Nadu, India.
**Dr.R.Gunavathi**
Head of the Department(MCA)
Sree Saraswathi Thyagaraja College
Pollachi, Tamil Nadu, India.

**Abstract- In the recent situations, the abduction of the children is becoming the great concern. Taking care of the children seems to be the greatest chore for the parents. Based on the reports produced, the kidnapping of the children for various reasons is reaching its hike. This paper gives the indication of the abduction and the measures to crack the problems.**

**Keywords- Child Abduction, Trafficking.**

## 1. INTRODUCTION

Abduction is defined as taking away a person by influence, by fraud, or by open force or violence. Child Abduction is the crime of wrongfully removing or wrongfully retaining, detaining or concealing a child or baby .

Child abuse is a state of emotional, physical, economic and sexual maltreatment meted out to a person below the age of eighteen and is a globally prevalent phenomenon which is found now a days[1].

According to World Health Organization (WHO), the term Child abuse is the term that comes with physical, emotional, sexual abuse.

- Physical Abuse: It describes that the children are made agonized physically like burning, punching, hitting, beating etc.
- Sexual Abuse: This means having inappropriate sexual behavior with the children.
- Emotional abuse: The children are affected psychologically, mentally due to the bizarre forms of punishment given by the parents/care takers
- Neglect: It is the failure to provide for the child's basic needs. Neglect can be physical, educational or emotional.

The abduction of children is of two types viz

- Family abduction
- Non Family abduction

The most common type of abduction is family abduction, which occurs when a child is taken by a parent or family member, where the event involves intent to deprive a lawful guardian of custodial privileges .Family abductions often occur because of dissatisfaction with custodial arrangements following a divorce, marital separation, or the breakup of a non marital relationship.

Nonfamily abduction is defined as an incident in which a stranger or non familial acquaintance takes or detains a child without lawful authority or permission from parents or legal guardian. The more dangerous type of nonfamily abduction is referred to as stereotypical abduction, or abduction perpetrated by a stranger or slight acquaintance, which occurs in conjunction with ransom, murder, or with the intent to keep the child permanently.

## 2. OBJECTIVE

The major goal of this study is

- To focus the causes for child abduction
- To understand the existing method of rescuing the children from the abduction.
- To apply the technology in rescuing the children from abductions.

## 3. THE CAUSES FOR ABDUCTION

It was analyzed that the ground modality of the abduction is called Trafficking. The protocol provided the definition of trafficking as, "the recruitment, transportation, transfer, harboring or receipt of persons by means of the threat or use of force or other forms of coercion, of abduction, of fraud, of deception, of the abuse of power or of a position of vulnerability or of the giving or receiving of payments or benefits to achieve the consent of a person having control over another person, for the purpose of abusement. Exploitation shall include, at a minimum, the exploitation of the prostitution of others or other forms of sexual abusement, forced labour or service, slavery or practices similar to slavery, servitude or the removal of organs"


Fig 1: Trafficking [2]

International Labor Organization (ILO) defines the child trafficking as "Child trafficking is about snatching the children out of their protective environment and preying on their weakness for the purpose of abusement".

## 4. CHILD ABDUCTION IN TAMILNADU

The total population of the children in Tamilnadu during 2016 is 1,38,20,661.

**Table 1.1: Projected children population of different age group in Tamilnadu**

| S.no | Age group in yrs | Male | Female | Total |
|---|---|---|---|---|
| 1 | 6-10 Yrs | 2848615 | 2686363 | 5534978 |
| 2 | 11-13 Yrs | 1931547 | 1775449 | 3706996 |
| 3 | 14-15 Yrs | 1255352 | 1148801 | 2404153 |
| 4 | 16-17 Yrs | 1132653 | 1041881 | 2174534 |

Among the children population in 2016, at least 2,741 children were kidnapped. In Tamil Nadu 16,183 children were reported missing between 2009-2014[3]. In the present scenario, the children in the age group of 4 to 15 are kidnapped for various reasons such as for illegal adoption, for sexual abuse, for removing the organs, for child labour and it becomes the major issue all over the world.

As per National Human Rights Commission (NHRC) report, during 2003 Tamilnadu takes $4^{th}$ place in trafficking and it was about 12.3% [2]. The report also produces that intra-state trafficking was the common phenomena in Tamilnadu. For each year, the abduction rate increases [4].

As per NHRC report, it was found that at least 2 children were missing every day, which may rise to trafficking. The trafficking might be for prostitution, handed over to criminal gangs or sold for illegal adoption [4].

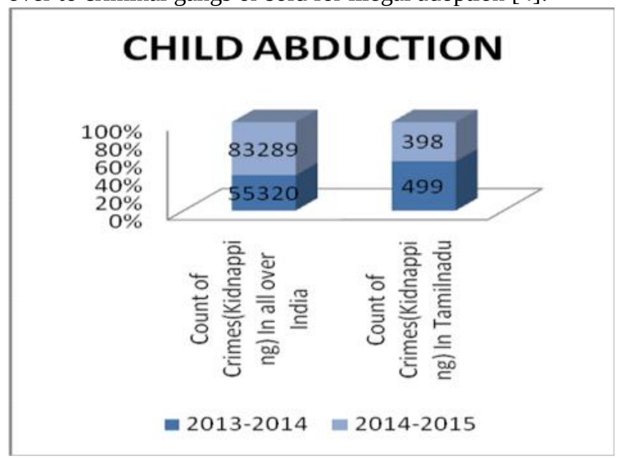



**Fig 2: Count of kidnapping**

## 5. THE EXISTING STRATEGY APPLIED TO RESCUE THE CHILDREN

There are diverse strategies are prevailing to rescue the children from kidnapping. Few of them are

- Getting help from an NGOs (Childline, Thozhamai)
- Getting help from Police.
- Through the tracking device

Through the Non Government Organizations like Childline, Thozhamai the services are being provided to the society by forming the committee and monitoring the children of the village or town. They are giving awareness programme to the children and as well as to their parents as the way to safeguard the children in all circumstances. Through their helpline also, they are providing the service to help the children.

One more measure which applied in rescuing the children from abduction is getting assistance from police. The Policemen are providing their service to the children community through various means. As on April 30 2015, according to the SCRB (state crime records bureau) records, 181 boy and 151 girl missing cases are now under investigation by the Chennai city police[5]. Through the CCTV footage the police officers are tracing the culprits who kidnapped the children.

Even though all the measures are taken by different NGOs and the police officers, the entire group involved in kidnapping is not getting rid of. They can safeguard only few (say approximately only 5%)children. The remaining was not yet found. This problem is becoming major issue. So, some measures are required to safeguard the children from abduction.

The Department of Police formed the Zip net network , a portal to file the details about the missing children in the year 2012. Unexpectedly it was found that the actual missing of persons were much more than the report submitted by NCRB. So, the Government of India has taken the steps to create the mechanism of forming Integrated Child Protection Scheme and Integrated Human Trafficking Units which comprises of the NGOs and Civil Society[6].

## 6. CONCLUSION

Thus, the existing scenario that is prevailing in Tamilnadu regarding Children abduction and the measures taken to preserve the children is in underneath level. To evade this kind of problem some technological measures is to be taken. In future, this kind of problem can be avoided using the technology called Wireless sensor network. The sensors can be built at affordable cost.

## REFERENCES

[1] Loveleen Kacker, Srinivas Varadan,Pravesh Kumar, “Study on Child Abuse INDIA 2007”, Kriti, New Delhi,2007.
[2] “Child Trafficking in India”, HAQ: Centre for Child Rights, 2016.

## NEWSPAPER

[3] Sibi Arasu, “9 kids go missing in Tamil Nadu every day. But rescuing them isn’t enough”, catch news, February 2017.
[4] Anuradha Nagaraj, “Missing children in Tamil Nadu raise trafficking fears”, Reuters India,April 1, 2016.
[5] A Subramani, “Hundreds of child missing cases in Tamil Nadu come under Madras high court scrutiny”, The times of India, Jun 17, 2016.
[6] Shakti Vahini, “Missing Children In India Government initiatives And Court Orders”, trackthemissingchild.gov.in/.../News.../NATIONAL_LEGAL_RESEARCH_DESK.pdf.

# A Survey on Opinion mining Techniques on Online Reviews

[1]Mrs. N. Sathyapriya, [2]Mrs. C. Akila

Assistant Professor
Department of Information Technology,
Sri G.V.G Visalakshi College for women, Udumalpet

***Abstract:*** **Today's world is a world of Internet. All work can be done with the help of Internet. Internet has become a new source of entertainment, education, communication and shopping etc. Sentiment Analysis is a subfield of Natural Language Processing concerned with the determination of opinion about a product or topic. Sentiment Analysis also called Opinion Mining, which collects and examines opinions about the product made in blog posts, comments or reviews. Sentiment Analysis is à rapidly growing area. There are numerous e-commerce sites available on Internet which allows users to give feedback about specific products. Users post their feedbacks as reviews in the websites which is helpful for companies and other users. These reviews are analyzed to determine the opinion of the users about the objects. It is impossible to manually analyze those reviews. To overcome the problem, many algorithms are proposed for mining the opinions of the users. Algorithms enable us to extract opinions from the Internet and predict customer's preferences. This paper presents various techniques used for opinion classification by different authors and its accuracy in the classification of opinions.**

***Index Terms*****-- Accuracy, Algorithms, Opinion, Opinion Mining, Reviews, Techniques, Sentiment Analysis.**

## I. INTRODUCTION

Opinion Mining is the field of study that analyzes people's opinion, sentiments, evaluations, attitude and emotion from written text. Opinion Mining is one of the most active research area in Natural Language Processing. Sentiment Analysis defines a process of extracting, identifying, analyzing and characterizing the sentiments or opinions in the form of textual information using Machine Learning, Natural Language Processing or Statistics. Today many companies are using Sentiment Analysis as the basis for developing their marketing strategies. Opinion Mining is the process of analyzing the text about a topic written in natural language and classify them as positive, negative or neutral based on the human sentiments, emotions, opinions expressed in it.

## II. LEVELS OF OPINION MINING

Opinion Mining can be done in three levels.

**Document Level -** In this, the whole document is considered as a single entity and the analysis is applied on the whole document and the result is summarized.

**Sentence Level -** In this, every sentence in the document is considered as an entity and the analysis is applied on individual sentence and the result is summarized to get the overall result of the document.

**Feature Level -** In this, object features are extracted from the comment and determines whether the stated opinion is positive or negative. It then groups the feature synonyms and produces a summary report.

## III. COMPONENTS OF OPINION MINING

The three main components of opinion mining are,

Opinion Holder - Person who expresses the opinion.

Opinion Object - Object on which opinion is given.

Opinion Orientation - Determines whether the opinion is positive, negative or neutral.

For example, "The picture quality of this mobile is good". In this review, the user who has written this review is the opinion holder. The opinion object is the "picture quality of the mobile" and the opinion is "good" which is a positive orientation.

## IV. Structural Design of Opinion Mining:

The opinion mining process involves three steps.

**Opinion Retrieval** - The process collects reviews about products from different sources and stores in the review database.

**Opinion Classification** - Opinion Classification is the key step in opinion analysis. Using sample reviews, the classifiers are trained and the trained classifier model is used to predict the category of new reviews.

**Opinion Summarization** - It sums up the overall positive and negative opinion expressed in the document. It involves the following steps,

- Extract all opinion terms.
- Classify the opinion terms as positive or negative.

If the number of positive terms exceeds the number of negative terms, the document is considered as positive opinion. If the reverse holds, the document is considered as negative opinion.

## V. Opinion Types

### I. Regular and Comparative Opinion

**Regular Opinion** - A regular opinion is often referred simply as an opinion in the literature and it has two subtypes.

- Direct Opinion
- In direct Opinion

**(*) Direct Opinion** - A direct opinion refers to an opinion expressed directly on an entity or an entity aspect. For example, "The sound quality of the mobile phone is good".

**(*) In direct Opinion** - It refers to an opinion that is expressed indirectly on an entity or an entity aspect. For example, “After taking this medicine, my joints felt worse”

**Comparative opinion -** A comparative opinion expresses a relation of similarities or differences between two or more entities. For example, the sentences, “Boost tastes better than Horlicks” and “Boost tastes the best” express two comparative opinions.

### II. Explicit and Implicit Opinion

**Explicit opinion -** An explicit opinion is a subjective statement that gives a regular or comparative opinion, e.g., “Coke tastes great,” and “Coke tastes better than Pepsi.”

**Implicit (or implied) opinion -** An implicit opinion is an objective statement that implies a regular or comparative opinion. Such an objective statement usually expresses a desirable or undesirable fact, e.g., “The sound quality of Nokia mobiles is longer than Samsung mobiles.”

## VI. Objective and Subjective Sentence

An objective sentence presents some factual information about the world, while a subjective sentence expresses some personal feelings, views, or beliefs. An example of objective sentence is “Smart Phone is an Apple product.” An example subjective sentence is “I like this mobile.”

## VII. Survey on Opinion Mining Techniques and its Accuracy

| S.No. | Month and Year of Publication | Authors | Title | Techniques used for Opinion Mining | Accuracy Produced |
|---|---|---|---|---|---|
| 1. | July, 2013 | Norlela Samsudin, Mazidah Puteh, Abdul Razah Hamdan, Mind Zakree, Ahmad Nazri | Immune Based Feature Selection for Opinion Mining | Naive Bayes<br>K Nearest Neighbor<br>Support Vector Machine | 91%<br>79%<br>92% |
| 2. | January, 2014 | Gaurangi Patel, Varsha Galande, Vedant Kekan, Kalpana Dange | Sentiment Analysis using SVM | SVM | N/A |
| 3. | February, 2014 | Pravesh kumàr Singh, Mohd Shahid Husain | Methodological Study of Opinion Mining and Sentiment Analysis | Naive Bayes<br>Multi Layer Perceptron | 76%<br>81% |

| | | | | | |
|---|---|---|---|---|---|
| | | | Techniques | Support Vector Machine | 81% |
| 4. | March, 2014 | R.Vinoth, Anjana Jayachandran, M.Balaji, R.Srinivâsan | A Hybrid Text Classification Approach using KNN and SVM | NN-SVM | N/A |
| 5. | April, 2014 | Gayathri Deepthi, K. Sashi Rekha | Opinion Mining and Classification of User Reviews in Social Media | Naive Bayes | 84% |
| 6. | April, 2014 | Nidhi R. Sharma, Prof. Vidya D Chitre | Opinion Mining, Analysis and its Challenges | Expectation Maximization based on Naive Bayes | N/A |
| 7. | May, 2014 | Richa Sharma, Shweta Nigam, Rekha Jain | Mining of Product Reviews at Aspect Level | Dictionary based Approach | 67% |
| 8. | October, 2014 | Vaishali Mehta, Prof. Ritesh K Shah | Approaches of Opinion Mining and Performance Analysis: A Survey | Naive Bayes<br>Genetic Algorithm<br>Naive Bayes+Genetic Algorithm | 91%<br>91%<br>93% |
| 9. | March, 2015 | Md.Safdar, Dr.Md.Jawed Ikbal Khan, Md. Daiyan | Mining Explicit Features for Opinion mining of Customer Reviews | Support Vector Machine+ Naïve Bayes | N/A |
| 10. | March, 2015 | Poobana.S, Sashi Rekha.K | Opinion Mining from Text Reviews using Machine Learning Algorithm | Support Vector Machine+Naïve Bayes | N/A |
| 11. | April, 2015 | Aamera Z.H.Khan, Dr.Mohammed Atique, Dr.V.M.Thakare | Sentiment Analysis using SVM | Support Vector Machine | N/A |
| 12. | June, 2015 | Gautami Tripathi, Naganna.S | Feature Selection and Classification approach for Sentiment Analysis | Support Vector Machine | 85% |
| 13. | August, 2015 | Gagandeeep Singh, Kamaljeet Kaur Mangat | Performance Analysis of Supervised Learning Methodologies for Sentiment Analysis of Tweets | Naive Bayes<br>Maximum Entropy<br>Support Vector Machine<br>K Nearest Neighbour | 71%<br>75%<br>81%<br>85% |
| 14. | August, 2015 | Dola Saha, Prajna Paramita Ray | Sentiment Analysis on Tweet Dataset using Datamining Techniques | K-means | 77% |
| 15. | September, 2015 | Preety, Sunny Dahiya | Sentiment Analysis using Support Vector Machine And Naïve Bayes Algorithm | Naïve Bayes<br>Support Vector Machine<br>Naïve Bayes+K Means | 80%<br>84%<br>89% |
| 16. | 2015 | K. Uma Maheswari, S.P. Raja Mohana, G. Aishwarya Lakshmi | Opinion Mining using Hybrid Methods | Support Vector Machine<br>Support Vector Machine+Particle Swarm Optimization | 68%<br>82% |
| 17. | 2015 | Chetashri Bhadane, Hardi Dalal, Heenal Doshi | Sentiment Analysis: Measuring Opinions | Support Vector Machine | 78% |
| 18. | February, 2016 | A.G.Dongre, Sushmit Dharurkar, Swanand Nagarkar, Rahul Shukla, Vivek Pandita | A Survey on Aspect based Opinion Mining from Product Reviews | Syntactic Based Approach – SentiwordNet | N/A |

## VIII. Conclusion

Opinion miming is an emerging field of data mining used to extract the pearl knowledge from huge volume of customer comments, feedback and reviews on any product or topic etc. A lot of work has been conducted to mine opinions in form of document, sentence and feature level sentiment analysis. Information from micro-blogs, blogs and forums as well as news source, is widely used in SA recently. This media information plays a great role in expressing people's feelings, or opinions about a certain topic or product. Using social network sites and micro-blogging sites as a source of data still needs deeper analysis.

This paper presented an overview on opinion mining sentiment analysis, its levels and types. The paper focused on the various opinion mining techniques applied by the authors and the accuracy produced in it.

**References:**

[1] Aamera Z.H.Khan, Dr.Mohammed Atique, Dr.V.M.Thakare, "Sentiment Analysis using SVM", International Journal of Advanced Research in Computer Science and Software Engineering, Vol.5, Issue.04, April 2015. (ISSN: 2277 - 128X).

[2] Ananta Arora, Chinmay Patil, Stevina Correia, "Opinion Mining: An Overview", International Journal of Advanced Research in Computer and Communication Engineering, Vol.4, Issue.11, November 2015. (ISSN: 2278-1021).

[3] G.Angulakshmi, Dr.R.ManickaChezian, "An Analysis on Opinion Mining: Techniques and Tools", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 3, Issue 7, July 2014.

[4] Asmita Dhokrat, Sunil Khilare, C. Namrata Mahender, "Review on Techniques and Tools used for Opinion Mining", International Journal of Computer Applications Technology and Research, Vol.4, Issue.6, 2015. (ISSN:2319 - 8656).

[5] Bing Liu, "Sentiment Analysis and Opinion Mining", Morgan and Claypool Publishers, 2012.

[6] Chetashri Bhadane, Hardi Dalal, Heenal Doshi, "Sentiment Analysis: Measuring Opinions", Science Direct, 2015.

[7] Dola Saha, Prajna Paramita Ray, "Sentiment Analysis on Tweet Dataset using Datamining Techniques", International Journal of Advanced Research in Computer Science and Software Engineering, Vol.5, Issue.08, August 2015.(ISSN: 2277 - 128X).

[8] A.G.Dongre, Sushmit Dharurkar, Swanand Nagarkar, Rahul Shukla, Vivek Pandita, "A Survey on Aspect based Opinion Mining from Product Reviews", International Journal of Innovative Research in Science, Engineering and Technology, Vol.5, Issue.2, February 2016. (ISSN: 2319 - 8753).

[9] Gagandeeep Singh, Kamaljeet Kaur Mangat, "Performance Analysis of Supervised Learning Methodologies for Sentiment Analysis of Tweets", International Journal of Advanced Research in Computer Science and Software Engineering, Vol.5, Issue.8, August 2015. (ISSN: 2277 - 128X).

[10] Gaurangi Patel, Varsha Galande, Vedant Kekan, Kalpana Dange, "Sentiment Analysis using SVM", International Journal of Innovative Research in Computer and Communication Engineering, Vol.2, Issue 1, January 2014. (ISSN: 2320 - 9801).

[11] Gautami Tripathi, Naganna.S, "Feature Selection and Classification approach for Sentiment Analysis", An International Journal, Vol.2, No.2, June 2015.

[12] Gayathri Deepthi, K. Sashi Rekha, "Opinion Mining and Classification of User Reviews in Social Media", International Journal of Advance Research in Computer Science and Management Studies, Vol.2, Issue.4, April 2014. (ISSN: 2321 - 7782).

[13] Kh airullah Khan, Baharum Baharudin, Aurnagzeb Khan, Ashraf Ullah, "Mining Opinion Components from Unstructured Reviews: A Review", Journal of King Saud University - Computer and Information Sciences, 26, 2014.

[14] Nidhi R. Sharma, Prof. Vidya D Chitre, "Opinion Mining, Analysis and its Challenges", International Journal of Innovations and Advancement in Computer Science, vol.3, Issue 1, April 2014. (ISSN: 2347 - 8616).

[15] Norlela Samsudin, Mazidah Puteh, Abdul Razah Hamdan, Mind Zakree, Ahmad Nazri, "Immune Based Feature Selection for Opinion Mining", Proceedings of the World Congress on Engineering, Vol.3, July 3-5, 2013. (ISSN:2078 - 0966).

[16] Pravesh kumàr Singh, Mohd Shahid Husain, "Methodological Study of Opinion Mining and Sentiment Analysis Techniques", International Journal on Soft Computing, Vol.5, No.1, February 2014.

[17] Poobana.S, Sashi Rekha.K, “Opinion Mining from Text Reviews using Machine Learning Algorithm”, International Journal of Innovative Research in Computer and Communication Engineering, Vol.3, Issue.3, March 2015. (ISSN: 2320 - 9801).

[18] Preety, Sunny Dahiya, “Sentiment Analysis using SVM and Naïve Bayes Algorithm”, International Journal of Computer science and Mobile Computing, Vol.4, Issue 9, September 2015. (ISSN: 2320 - 088X).

[19] Reema Verma, Dr.Kiranjyoti, "Opinion Mining and Analysis of the Techniques for User Generated Content", International Journal of Advanced Research in Computer Science and Software Engineering, Vol.5, Issue 5, May 2015. (ISSN:2277 - 128X).

[20] Richa Sharma, Shweta Nigam, Rekha Jain, “Mining of Product Reviews at Aspect Level”, International Journal in Foundations of Computer Science & Technology, Vol.4, No.3, May 2014.

[21] Rushabh Shah, Bhoomit Patel, “Procedure of Opinion Mining and Sentiment Analysis: A Study ”, International Journal of Current Engineering and Technology, Vol.4, No.6, December 2014.

[22] Md.Safdar, Dr.Md.Jawed Ikbal Khan, Md. Daiyan, “Mining Explicit Features for Opinion mining of Customer Reviews”, International Journal of Emerging Technology and Advanced Engineering, Vol.5, Issue 3, March 2015. (ISSN: 2250 - 2459).

[23] G. Sneha, CT.Vidhya, "Algorithms for Opinion Mining and Sentiment Analysis: An Overview", International Journal of Advanced Research in Computer Science and Software Engineering, Vol.6, Issue.02, February 2016. (ISSN: 2277 - 128X).

[24] K. Uma Maheswari, S.P. Raja Mohana, G. Aishwarya Lakshmi, “Opinion Mining using Hybrid Methods”, International Journal of Computer Applications, 2015. (ISSN: 0975 - 8887).

[25] Vaishali Mehta, Prof. Ritesh K Shah, “Approaches of Opinion Mining and Performance Analysis: A Survey”, International Journal of Advanced Research in Computer Science and Software Engineering, Vol.4, Issue.10, October 2014. (ISSN: 2277 - 128X).

[26] Vidisha M. Pradhan, Jay Vala, Prem Balani, “A Survey on Sentiment Analysis Algorithms for Opinion Mining”, International Journal of Computer Applications, Vol.133, No.9, January 2016. (ISSN: 0975 - 8887).

[27] R.Vinoth, Anjana Jayachandran, M.Balaji, R.Srinivâsan, ” A Hybrid Text Classification Approach using KNN and SVM”, International Journal of Advance Foundation and Research in Computer, Vol.1, Issue.3, March 2014.(ISSN:2348 - 4853).

[28] Walaa Medhat, Ahmed Hassan, Hoda Korashy, “Sentiment Analysis: A Survey”, Ain shams Engineering Journal, 5, 2014. (1093 - 1113).