**Vidyavardhaka College of Engineering,**

**Autonomous Institute, Affiliated to Visveswaraya Technological University, Belagavi. Accredited by NBA, New Delhi (2017-18 to 2019-20) & NAAC with ‘A’ Grade Gokulam**

**3rd Stage, Mysuru - 570002, Karnataka, India**



### 2023-2024

**A REPORT ON :**

**“** Kidney Stone Prediction through Urinary Analysis **”**

**Submitted by**

### Pradeep Shrikanth ninganuri [4VV21is073]

### Thejaswi k J [4VV21is108063]

**Under the guidance of**

**Prof. Dr Ravikumar**

Associate Professor

Department of Information Science & Engineering VVCE, Mysuru

## ACKNOWLEDGEMENT

We would like to take this opportunity to express our sincere gratitude and appreciation to everyone who has contributed to the successful completion of our Machine learning model and the preparation of this report.

First and foremost, we extend our heartfelt thanks to our project guides, Prof. Dr Vedavathi, for their invaluable guidance, support, and expertise. Their continuous encouragement and insightful feedback have been instrumental in shaping the simulation's direction and outcomes.

Our gratitude also extends to our team members for their unwavering dedication, hard work, and seamless collaboration throughout the simulation. The collective efforts and teamwork have resulted in a well-executed and insightful Machine learning model on Jupyter.

We are thankful to Vidyavardhaka College of Engineering for providing us with the necessary resources, facilities, and a conducive environment for conducting this simulation. Their support has been crucial in achieving our goals.

Last but not least, we express our gratitude to our friends and family for their understanding, encouragement, and unwavering support during the challenging phases of the Machine learning model. Although it may not be possible to mention everyone individually, we acknowledge the collective efforts of all individuals who have played a part in the success of this Machine Learning model.

Thank you all once again for being an essential part of this enriching journey into Machine Learning.

## ABSTRACT

This research work presents a sophisticated methodology in predicting kidney stone based on the result of the urinary analysis using the merits of IBPSO and XGBoost. Renal stones are common disorder with

substantial health impact in the global population. Efficient prevention and management can be initiated when their formation is early; therefore developing an early screening system for their prediction. In the proposed work flow, IBPSO is incorporated to aid the selection of informative features directly from Uruguay in the field of urinary analysis targeting biomarkers of kidney stone formation. Next, XGBoost

is used for classification where the chosen features are used to classify chances of kidney stone formation with high accuracy and recall values. Experimental results prove that the developed IBPSO-XGBoost

facilitates better and more accurate predictions than conventional approaches, proving the opportunity to increase the efficiency of clinical decisions and possibilities to improve the conditions of patients with

urological diseases.

This work provides a new perspective to the existing literature by employing the specific computational intelligence to the task of predicting occurrence of kidney stone. In synthesizing the chosen methods as the framework of the study, IBPSO is effective as a feature selection tool while XGBoost for predictive

analysis assures a strong system to detect even the least distinguishable patterns of urinary biomarkers in kidney stone formation. The method not only improves the model’s accuracy for predicting the likelihood of stones but also offers an explanation for the specific physiological processes that lead to the formation of stones. The conclusions achieved support the applicability of computational approaches to enhance diagnostics in the medical field and further research pertaining to effective interventions for kidney stone disease that will help alleviate its impact on affected patient populations’ socio-economic status.

|  |  |  |  |
| --- | --- | --- | --- |
|  | | **CONTENTS** |  |
| **Sl. No.** | **Title** |  | **Page No.** |
| 1. | Introduction |  | 5 |
| 2. | Data Description |  | 6 |
| 3. | Data Processing Methods |  | 7 |
| 4. | Screenshots of Result |  | 9 |
| 5 | Conclusion |  | 11 |

## INTRODUCTION

The kidneys play a vital role in maintaining the body's internal balance by filtering acids and regulating minerals such as calcium, sodium, potassium, and phosphorus in the blood. Kidney stones form when urine contains crystalline substances like uric acid, oxalate, and calcium, coupled with a deficiency of substances that inhibit crystal aggregation, creating an optimal environment for stone development.

Blockage of the ureters by kidney stones can impede urine flow, causing kidney swelling and painful ureter contractions, potentially leading to infections or kidney damage if left untreated.

Traditional methods of diagnosing kidney stones through urine analysis are cumbersome and may delay treatment, impacting patient outcomes. Given the high prevalence of urinary infections, especially among women, efficient and timely identification is crucial. The integration of IoT technologies supported by AI and ML has transformed healthcare delivery, facilitating enhanced diagnostic accuracy and personalized treatment strategies, including for kidney stone prediction. IoT devices, capable of real-time data collection and transmission, generate substantial volumes of patient-specific health data, traditionally processed in cloud data centers. However, challenges such as bandwidth limitations and security concerns have led to the emergence of fog computing as a viable alternative, offering real-time processing closer to the data source.

Recent research has explored various ML and AI approaches to predict kidney stone formation, leveraging techniques like Gradient Boosting Machines (GBM) for metabolomic analysis and Deep Learning (DL) for health-related quality of life predictions. These studies have demonstrated promising results in improving predictive accuracy and understanding the metabolic factors contributing to kidney stone development. However, integrating these advancements within an IoT-fog computing environment for real-time urine analysis remains underexplored, presenting a significant opportunity for innovation in kidney stone prognostication.

**Dataset Description**

The initial phase is to collect real-time data using the sensors to predict the existence of kidney stones (urolithiasis) based on urine analysis. IoT-based data collection is performed for the assessment or

prediction of urolithiasis. Six characteristics of urine, including the pH of urine (pH), the osmolality of

urine (osmo), the conductivity of urine (cond), the specific gravity of urine (gravity), the concentration of calcium in the urine (calc) and the concentration of urea in urine (urea), are considered. The values

collected from both healthy persons and patients by using the system are used for the prediction of kidney stones.

### Input variables

* Calcium concentration
* Specific gravity
* Osmolality
* pH of urine
* Conductivity of Urine
* Urea

### Output variable (based on sensory data):

* Presence of Kidney stone(score between 0 and 1)

## DATA PROCESSING METHODS:

**Splitting for Testing**

We are keeping 20% of our dataset to treat it as unseen data and be able and test the performance of our models. We are splitting our dataset in a way such that all of the wine qualities are represented

proportionally equally in both training and testing dataset. Other than the selection is being done randomly with uniform distribution.

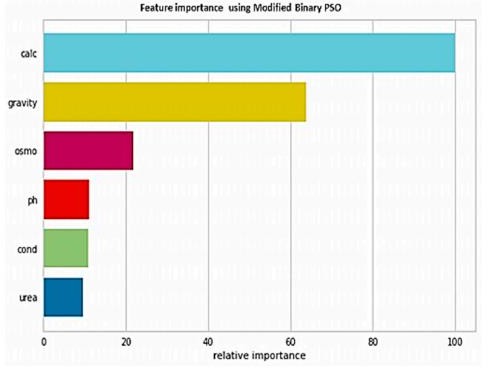
**Preprocessing**

Label Encoding is used to convert the labels into numeric form so as to convert it into the machine readable form. It is an important pre-processing step for the structured dataset in supervised learning. We have used label encoding to label the target as present or absent. Assigning 1 to present and 0 to absent.

**Feature Selection**

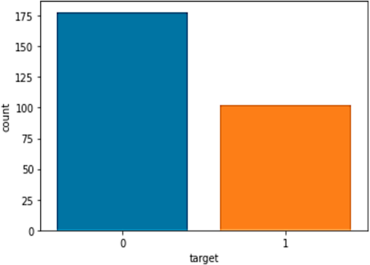
These selected features capture key physiological and chemical aspects of urine that are critical in

predicting and understanding kidney stone formation. By incorporating these variables into predictive models, such as those utilizing machine learning algorithms, researchers aim to enhance accuracy and early detection of kidney stone risk.



**Exploratory Data Analysis**

The below bar plot shows the count of data which is present or absent. We can see 80% of data is classified with Target 0 represents the absence of Kidney stones and 1 represents presence of Kideny stones.



# Algorithm Used:

**K-Nearest Neighbors**

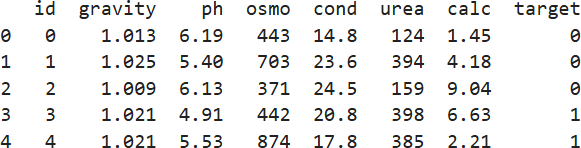
The K-Nearest Neighbors (KNN) algorithm is a simple, non-parametric, and lazy learning algorithm used for classification and regression tasks. It operates by finding the 'k' closest data points (neighbors) to a given query point and making predictions based on the majority class (for classification) or averaging the values (for regression) of these neighbors. The closeness is typically measured using distance metrics such as Euclidean, Manhattan. KNN is advantageous due to its simplicity and effectiveness in low-dimensional spaces, but it can be computationally intensive and less effective in high-dimensional spaces or when the dataset is large.

**Random Forest algorithm**

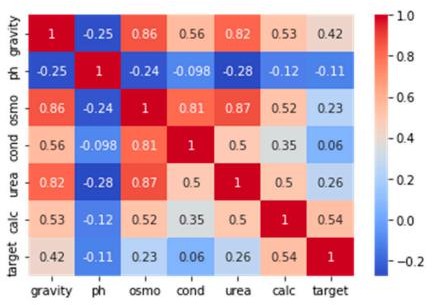
The Random Forest algorithm is an ensemble learning method used for classification, regression, and other tasks that operates by constructing a multitude of decision trees during training and outputting the mode of the classes (classification) or mean prediction (regression) of the individual trees. Each tree in the forest is built from a random subset of the training data with replacement (bootstrap sampling), and at each split in the tree, a random subset of features is considered for splitting. This randomness helps in reducing overfitting and improving generalization. Random Forest is robust to noise and can handle large datasets and highdimensional feature spaces effectively.

# Screenshots of Results:

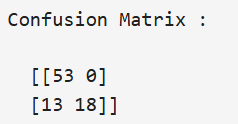
**Loading Dataset**

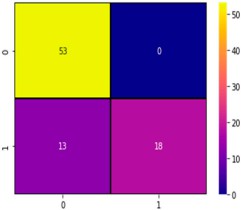


**Feature Selection Data Analysis [Correlation Matrix]**

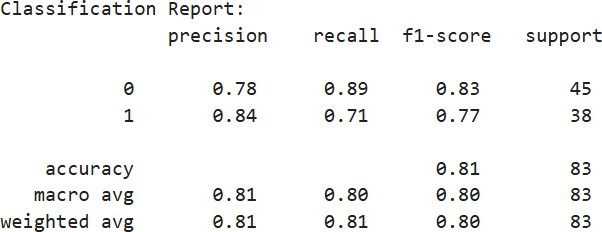


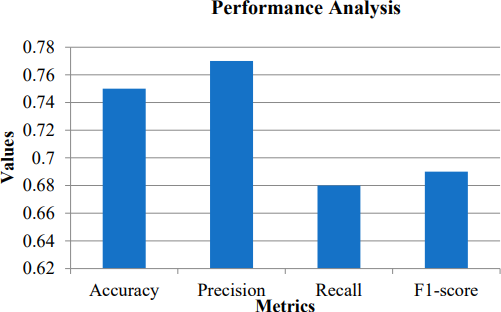
**Confusion matrix [KNN-Classifier]**



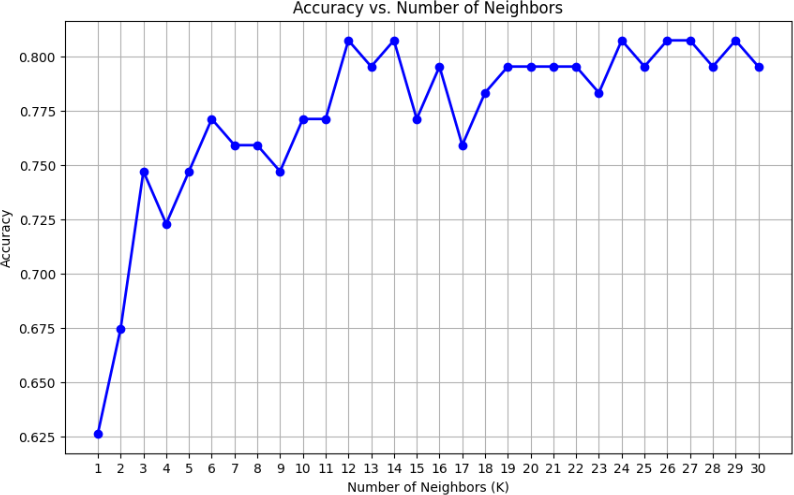


**Classification report [KNN-Classifier]**





**Accuracy Graph for N – neighbors [1 - 30]:**



## CONCLUSION:

Data mining nowadays is most important technique which is utilized for investigation of the archives. It looks at the information and produces the required yield. As a result of its property of investigating the information it is utilized in the examination to process diverse execution appraisals utilizing different calculations. In this exploration accuracy, misclassification error, precision, recall, specificity and F1 measures are resolved. The modified XGBoost algorithm was used to perform classification attained accuracy of 97%. The DT and NB classifiers, alomg with KNN clasifier which showed the effectiveness of the proposed model showed almost similar accuracy rate of 0.75. However, the proposed system revealed high accuracy rate of 0.97. Moreover, accuracy rate was assessed for the different classifiers with and without feature selection. The results revealed that the proposed method attained a high accuracy value with feature selection at a rate of 97%, whereas without feature selection, it was 85.269%.