**Project Name: Under-Five Mortality Rate (Per 1000 Live Births)**

**Introduction:**

Our Project is to find out the under-five mortality rate for per 1000 live births. These days, the highest death rate of children in the first 28 days is the neonatal period. It's very difficult to have safe childbirth and effective care during the neonatal period because the majority of deaths occur in children under 5 years old and the 44% of the deaths of children during the neonatal period [1]. The most causes of the death during this phase are preterm birth and compilation during birth for example infarction and breathing cause. When children grow and clear the neonatal period then the first to 5 years is most difficult for the children. After the neonatal period the main causes for the death of the children is shifting to diseases like pneumonia, diarrhoea and malaria [1].

According to the research, every year nearly 3 million children die in their first month and a similar number stillborn. During the first month, around 25% died in the first week and the remaining 75% died in the first week. After the child, the first 48 hours are very important for the child's survival. That’s why both mother and child need follow-up care during this time.

Those children who are under the age of 5, face a big risk of dying. Actually, this problem is very bad in Sub-Saharan Africa and Southern Asia. In these areas most of the cases of child death are under the age of 5 years. On the other hand, for example in Europe and America the death ratio of the child is going low [1].

**Study Review:**

Our main Project is to identify the mortality rate under 5 years for per 1000 live births. In the previous research, many researchers applied machine learning and Deep Learning Models. Actually, in machine learning, they have used the Linear Regression, Support Vector Machine, Decision Tree, Random Forest and Advance Methods like (Assembling Techniques). In Deep Learning, they have used the Neural Network for Example (FCN with ML Model). Before Applying the Machine Learning Models. They have splatted the Dataset into 70% Training and 30% Testing then they have applied the Machine Learning Models. In the many research papers Random Forest quite performs well as compared to other ML Models [2].

Actually, Random Forest is the Ensemble Technique and it's a combination of multiple Decision Trees. Actually, In the Random Forest Tree the error rate is very less due to ensemble methods. Actually, In the Random Forest Tree they send his error in the next step so that the next step will learn this error and modify it. That’s why this method performs well in many papers.

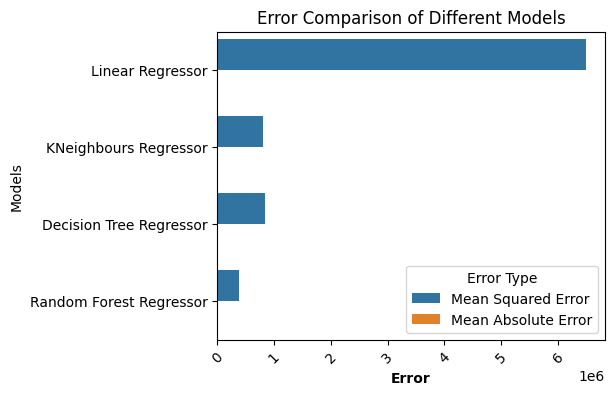
In the Research, the main evaluation errors such as (Mean Squared Error and Mean Absolute Error). Actually, these evaluation errors are used for the selection of the Machine Learning Models [3]. We’ve multiple errors for the selection of the Machine Learning Models but (Mean Squared Error and Mean Absolute Errors) are used in many papers.

**Methodology:**

Our problem is a regression problem that’s why I’ve used the Machine Learning Supervised Learning Regression Models such as (Linear Regression, Decision Tree, Random Forest and K Neighbours). Now, I’m talking about the steps of my project [3].

First Step, I’ve cleaned the Dataset. In this step, I’ve converted the Categorical Data to Numerical Data using Label Encoder and Normalised the Numerical Data using Limescale. Second Step, I’ve divided the Dataset into 80% Training and 20% Testing. Third Step, I’ve applied the 4 main Machine Learning Models like (Decision Tree, Random Forest, Linear Regression and K Neighbours). Now, in the paragraph below I’m talking about the models or the evaluation of the Model [4].

Linear Regression Model, First I’ve trained the Linear Regression Model in the 80% of Dataset and After that I’ve tested the Linear Regression Model in the 20% Dataset then I’ve evaluated this model using Mean Squared Error and Mean Absolute Error. The Mean Squared Error of the Linear Regression is 6495313.09 and Mean Absolute Error is 2137.188. In the Decision Tree Model, Firstly, I trained the Decision Tree Model in the 80% of the Dataset and After that I tested the Decision Tree Model in the 20% Dataset then I evaluated this model using Mean Squared Error and Mean Absolute Error. The Mean Squared Error of the Decision Tree is 841414.69 and Mean Absolute Error is 232.05. In the K Neighbours Model, Firstly, I trained the K Neighbours Model in the 80% of the Dataset and After that I tested the K Neighbours Model in the 20% Dataset then I evaluated this model using Mean Squared Error and Mean Absolute Error. The Mean Squared Error of the K Neighbours is 2277868.51 and Mean Absolute Error is 672.95. In the Random Forest Model, Firstly, I trained the Random Forest in the 80% of the Dataset and After that I tested the Random Forest Model in the 20% Dataset then I evaluated this model using Mean Squared Error and Mean Absolute Error. The Mean Squared Error of the Random Forest is 411605.38 and Mean Absolute Error is 187.73[4]. In the Below Figure, I’ve displayed the Mean Squared Error and Mean Absolute Error using the graph.



In the Figure, Random Forest has the least error as compared to other models.

**Conclusion:**

In this Project, I’ve studied the under-five mortality rate per 1000 live births, finding the challenges faced during the neonatal period and the subsequent years up to age five. We found that the highest mortality rates occur within the first month of life, with preterm birth and birth complications being major contributors to neonatal deaths [5]. Beyond the neonatal period, diseases such as pneumonia, diarrhoea and malaria pose significant threat to child survival. In my research, my aim is to predict the under-five mortality rates using the Machine Learning Models and I’ve applied the Decision Tree, Random Forest, Linear Regression and K Neighbours. But Random Forest performs well as compared to the other Models That’s why Random Forest easily can find out the relationship within the data.

**Summary:**

In my research, I’ve predicted the under-five mortality rates, which remain a critical global health concern, particularly in regions like Sub Saharan Africa and Southern Asia [6]. In this research I’ve mentioned the challenges during the neonatal period and the subsequent years up to five years. In my whole project I’ve used machine learning such as (Linear Regression, Decision Tree, Random Forest and K Neighbours) and Random Forest Model performing good instead of Other Models and Random Forest is quite efficient to predict the under-five mortality rates.

**References:**

[1] Silva, R. (2012). Child mortality estimation: consistency of under-five mortality rate estimates using full birth histories and summary birth histories.

[2] Khodaei, G. H., Khademi, G., & Saeidi, M. (2015). Under-five Mortality in the World (1900-2015). *International Journal of Paediatrics*, *3*(6.1), 1093-1095.

[3] Raheim, M. H., & Masih, M. (2021). A Descriptive Study of Under-five Mortality. *Journal of Pharmaceutical Research International*, 86-97.

[4] Kipp, A. M., Blevins, M., Haley, C. A., Mwinga, K., Abimana, P., Shepherd, B. E., ... & Vermund, S. H. (2016). Factors associated with declining under-five mortality rates from 2000 to 2013: an ecological analysis of 46 African countries. *BMJ open*, *6*(1), e007675.

[5] Gebretsadik, S., & Areophane’s, E. (2016). Determinants of under-five mortality in high mortality regions of Ethiopia: an analysis of the 2011 Ethiopia Demographic and Health Survey data. *International Journal of Population Research*, *2016*.

[6] Rahmati, P. D., Kurnia, I. D., Asih, M. N., Muriwai, T. W., Krishana, I., Arief, Y. S., ... & Arifin, H. (2022). Determinants of under-five mortality in Indonesia: A nationwide study. *Journal of Paediatric Nursing*, *65*, e43-e48.