**Knowledge Discovery from Diabetes in Children**

* 1. **References**

Aboul Ella Hassanien, Mohamed E. Abdelhafez, and Hala S. Own. Rough Sets Data Analysis in Knowledge Discovery: A Case of Kuwaiti Diabetic children Patients

**1.2 Purpose of study**

One of the significant issues in the present period is the prevalence of diabetes among children and adolescents. The majority of research demonstrates how this condition affects children's lives. It has a significant effect and leads to emotional and behavioral issues. Previous data analysis algorithms gathered the majority of the medical data through assumptions. The outcomes of such data are therefore inaccurate. Hence In this essay, the author uses the instance of Kuwaiti diabetic children's patients to discuss rough sets data analysis and knowledge discovery.

**1.3 Research design and strategy**

There are two main phases to the examination of rough set data.

**Preprocessing stage:**

Pre-processing involves cleaning the data and discretizing the characteristics before the data sets are used. Data is divided into test and training sets, and classification is approved. Decisions are provided if the categorization technique is successful. Preprocessing stage involves process and tools, and it is most important step for data understanding and had great impact on results.

**Processing Stage:**

The final step of creating rules is the processing phase.

**Relevant attribute extraction and Reduction:**

The data that is obtained from the medical industry includes a lot of redundant qualities that need to be categorized. Therefore, at this stage of processing, all the crucial traits will be found, and the redundant ones removed. And missing values are removed. It is referred to as attribute reduction. In some models methods like principal component analysis are used for dimensionality reduction. For filling missing values aggregate functions like mean, median mode are used.

**Rule Generation and Classification:**

The classification of these qualities is utilized to generate the decision rules. Combining the values of the characteristics results in these rules. This produced rule aid in correctly identifying the data. Based on the generated rules the target data is identified.

**1.4 Conclusion**

In the realm of medicine, gathering crucial, non-redundant data is challenging and fraught with ambiguity. The data mostly is raw format and problems like redundancy, missing values. The author provided a solution to data analysis problem using rough set theory has been used to solve this issue.

**1.5 Contribution**

The essay provided a thorough explanation of the idea of rough sets and demonstrated its applicability to the classification of medical data using a case study. The linked publications that are covered in the paper examine several cutting-edge methods for handling this uncertainty, including the use of neural networks, Bayesian classifiers, fuzzy theory, and more. Other theories that aid in establishing crucial medical norms and information include case-based reasoning and decision trees.

**2.1 Overall Assessment**

The author's detailed explanations made the material easy to read. In addition, the author did a great job of keeping a pleasant flow while offering clear explanation. The paper's primary merit is its explanation of underlying algebra and analysis. I must commend writers' writing abilities.

**2.2 Research methodology**

The author clearly explained rough sets data analysis scheme.

The Data set is passed to data pre-processing stage and data is cleaned and passed to data completion stage after which the data is sent to data discretization and then processing phases. After this the output data is split into training data set and testing data set. From the training data set we need to extract relevant attributes and perform attribute reduction and rule generation process. Generated rules are passed to the classifier. Finally with the help of testing data and classification is done and model is evaluated by classification accuracy. If the accuracy is satisfied, then decision is taken. If accuracy is not up to the mark or not satisfied, then gain we need to restart from pre-processing phase. The process continues until we got estimated accuracy

**2.3 New Knowledge Learned**

I get to understand how the rough sets theory may be utilized to extract crucial information and categorize medical data. and how to apply them to real-world issues.

**2.4 Future Research**

Rough sets theory was the only tool used to create this case study. However, this may be further altered by applying a variety of cutting-edge approaches, such as fuzzy algorithms and neural networks.

**3.Questions to discuss**

1. How rough sets deal with limited labeled property of big data?
2. How rough sets deal with computational inefficiency and overfitting in attribute reduction?