<u>Homework Assignment 4</u>

"Prescriptive Analytics for Recommendation-Based Business Process Optimization"

These days digitalization and new sensor technologies are helping businesses to improve and continuously grow. They are dealing with a huge amount of data daily yet, they are not able to use the valuable information inside the data due to a lack of use of prescriptive techniques, by not integrating process data.

So, in order to overcome these issues, the data-mining-driven concept comes into play for continuously optimizing business processes, where we can take the help of metrics.

Metrics-based goal definition (R1) should be supported, and metric prediction should be facilitated, and recommendation generation should be proactive during process execution (R2). Along with that recommendations should comprise multiple actions (R5) in order to achieve the goal.

It takes two major components into account - real-time prediction and recommendation generation.

Real-time prediction consists of three steps, which are definition of the data basis, mining model generation and mining model application. So, for the definition of the data basis, we refer to data about completed process instances in the PWH. For that, we would be needing the attributes from the completed process and need to predict the value of the metric. An appropriate data mining method must be devised that employs the customized database as training data for mining model creation. An expert user should be given the ability to view the created mining model so that he can understand the forecast and adjust the parameters. The interpretability of the produced data is therefore model need to be high in comparison. Because binary trees are easier to grasp, we employ decision tree induction as a classification strategy.

Recommendation Generation- In a process instance that has been currently running, suggestion generation determines a suggested course of action. Several action items associated with process characteristics and goal values for each of them make up an action suggestion. We build decision trees and correlate the metric value's classification as a class label with a subset of the characteristics of a subset of the process instances. Then, each branch of the tree that leads first from the root node to a leaf node indicates a hypothetical decision rule for a suggestion with the label "OK". Definition of the data basis, mining model generation, mining model analysis, and recommendation processing are the four sequential phases that makeup recommendation generation.

We often believe that only qualities indicating influencing variables, such as machine settings, are taken into account when choosing attributes like using a predefined filter.

Mining Model Generation We distinguish between binary trees and n-ary trees in terms of the decision tree's structure; the former have precisely two child nodes per node, whilst the latter have arbitrary numbers. On the one hand, if we suppose that a typical process characteristic has more than two values, an n-ary tree exposes a bigger number of rules owing to the n-ary split. This makes mining model analysis more difficult. However, any decision rule in an n-ary tree may be supported if the maximum number of underlying process instances is smaller than a binary tree rule, and Height is constant. In other words, because the rules are formed from fewer process instances, they are intended to be less important. To further limit the trees, we opt to only use branches with equal and unequal interactions to accelerate tree induction and make suggestions simpler. Thus, there aren't any subsets limits placed on the branches. Recommendations become more flexible and generic as a result. We set a maximum height for the decision tree in terms of its height. Depending on the specific process and the number of process characteristics limit the amount of recommendations' action items.

Mining Model Analysis and Recommendation Processing There are four guidelines for the rule assessment: - Misclassification rate q: The proportion of process instances for which the class labeled is assigned to that instance does not clearly categorize; hence, the lower the misclassification rate, the more reliable the system. Percentage of underlying instances r: It is the number of instances we obtained from all instances when the complete tree was used as a training set. If r is less than 0.1, the minimum percentage is not met, therefore the rule is not applicable. The length of rule l: This rule is determined by the quantity of attribute-value combinations. The rule becomes simpler the smaller it is. Compliance with planned values c: It determines whether the suggested tool is being utilized. Compliance is defined as the percentage of attribute-value combinations that match; the higher the percentage, the faster and easier the process is. The defined r and q thresholds must be filtered out of all the rules to choose the best one, and the t score must then be computed for each rule. This score, which runs from 0-100, is derived from the above four sub-scores using a weighted aggregate based on business standards. In this article, the authors discussed prescriptive analytics for real-time, recommendation-based business process improvement. The proof of concept supports the approach's technological viability and effectiveness and highlights the significance of robust data-collecting infrastructures, particularly in manufacturing processes, to enable real-time process optimization.