Warsaw, dn. 10 września 2023 roku

mgr inż. Dawid Szpadzik
(professional title,name)

Wydział Chemiczny / Katedra Technologii Chemicznej
(Faculty/Unit)

Politechnika Warszawska
(University)

ABSTRACT OF THE DISSERTATION

„Study of the possibility of using statistical tools in qualitative analyses in production areas"

dissertation supervisor**: prof. dr hab. inż. Wioletta Raróg-Pilecka**

This dissertation presents the creation of statistical tools in quality control analysis, in household chemicals manufacturing industry, using standard software and an office computer.

Machine learning and examples of its application in the field of chemical engineering are described including a historical overview. Commonly used binary classification models such as logistic regression, random decision forest, classification trees and others are presented. The method used in this study (naive Bayes classifier algorithm) is described in more detail and illustrated with an example. Ways of evaluating the performance of binary classifiers using a confusion table are cited.

Due to the implementation nature of the research, the process of manufacturing housekeeping chemicals was described. It consisted of manufacturing liquid cleaning products based on aqueous solutions in a batch model. Attention was paid to quality management at the manufacturing plant (ISO standards), the purpose of which was to ensure that the products placed on the market comply with the manufacturer's declarations and applicable legal standards. Consideration was given to the financial aspect (cost intensity) of maintaining a physicochemical laboratory, which consisted of expenditures for consumables (e.g. reagents, pipettes, gloves) and with the employment of qualified personnel. The research work was undertaken as part of efforts to continuously improve the production process of household chemicals and optimize manufacturing costs in chemical industry. They sought to ensure adequate product margins and a competitive price for the consumer. These activities were particularly important in the context of inflationary pressures caused by the COVID-19 pandemic and the geopolitical situation in Europe.

In the experimental part, the first phase analyzed the manufacturing process carried out at the cooperating entrepreneur. The requirements for the tool under development and the success criteria were established. A research criterion (the classifier tool achieves accuracy better than random) and an implementation criterion (class assignment accuracy of at least 95%) were specified. The operation of the created application was described, as well as its functions aimed at fulfilling the requirements set in the preceding stage. The classifier is equipped with a number of parameters that can be modified by the user.

The second stage of the experimental part was to optimize the performance of the machine learning algorithm (known as training). First, a general benchmark - reference settings - was established. Then each parameter of the algorithm was iteratively modified to study the characteristics of its effect on the quality of classification. In the next step, all arguments were set off simultaneously, using those settings for which the best values of accuracy and correlation coefficient were previously obtained. The final step was to investigate the possibility of reducing the number of physicochemical analyses that use disposable materials and reagents.

It was proved that it is possible to develop a tool using a machine learning model, working on a personal computer, created using Microsoft Office software. It was confirmed that it is possible to use statistical tools in quality control analysis, in production areas (the research goal of the doctoral thesis). The accuracy of the developed classifier tool was greater than 50% - thus better than random assignment. The level of accuracy required for implementation was achieved (a minimum of 95%), while maintaining the criterion for assigning classes to samples from the training set (a minimum of 80%). In the test population, the lowest accuracy achieved was 97.85%. The optimization work showed that the number of cost-intensive physical-chemical tests can be reduced, which can further improve the safety of laboratory personnel.

The research work showed that mathematical corrections did not have the desired effect. Leveling the effect of the absence of occurrences of observations (Laplace smoothing) resulted in a significant deterioration of classification quality indicators. Also, no significant changes in metrics were observed when using standard deviation-based discretization of real variables.

The presented dissertation proved that it is possible to build a machine learning algorithm using standard software (Microsoft Office) and an office computer. The application met the requirements to achieve implementation potential in a manufacturing plant of Reckitt Benckiser Production (Poland). Classification accuracy reached 99.85%, and the correlation coefficient reached 0.9758. The parameters controlling the algorithm and their impact on the effectiveness of class determination were discussed. In summary, the capabilities of machine

learning tools were learned and proved that their application does not require specialized computer tools and can be used in standard manufacturing processes of liquids household products, reducing the costs associated with maintaining a quality control laboratory. Thus, the optimization of the activities of chemical engineers expands to incorporate mathematical modeling of the entire range of work to be performed, rather than merely on isolated physicochemical methods.

Dawid Szpadzik

mgr inż. Dawid Szpadzik

Keywords: machine learning, naive Bayes classifier, binary classification, chemical industry, quality control in chemical industry, production technology of household chemicals