|  |  |  |
| --- | --- | --- |
| **Title** | **:** | A Study on whether Automated Machine Learning or Generative AI can replace the Human Data Scientist |
| **Author** | **:** | Ryan Kho Yuen Thian |

**Abstract**

The rapid advancements in Artificial Intelligence (AI) and Automated Machine Learning (AutoML) have sparked a growing debate on whether these technologies can effectively replace human Data Scientists in the process of developing and implementing machine learning models. With automated tools becoming more sophisticated, the question of whether human expertise is still necessary is increasingly relevant. This study aims to explore the potential of Generative AI (represented by ChatGPT Data Analyst) and AutoML (represented by H2O AutoML) in executing machine learning tasks that are traditionally handled by human Data Scientists. The focus is on assessing whether these technologies can perform these tasks at a comparable level in terms of accuracy, efficiency, and adaptability. To evaluate these approaches, we use a credit risk assessment case study, with the German Credit dataset serving as the benchmark. Each of the three approaches—human Data Scientist, H2O AutoML, and ChatGPT Data Analyst—follows the CRISP-DM (Cross Industry Standard Process for Data Mining) methodology to build, evaluate, and deploy predictive models. By adhering to established Machine Learning Best Practices, this case study ensures a fair comparison across all methods. The strengths and weaknesses of each approach are critically analyzed, focusing on key factors such as capability, efficiency, reliability and flexibility. The human Data Scientist brings a deep understanding of domain-specific knowledge and customization, while AutoML and Generative AI aim to streamline the process, potentially reducing human involvement. The results of this study provide insights into whether Generative AI and AutoML can replace or complement human Data Scientists. This study also explores the possibility that a hybrid approach, combining human expertise with Generative AI and AutoML technologies, may offer the most promising solution.