Paper Title:
Ask Your Data—Supporting Data Science Processes by Combining AutoML and Conversational Interfaces

Paper Link:
Ask Your Data—Supporting Data Science Processes by Combining AutoML and Conversational Interfaces | IEEE Journals & Magazine | IEEE Xplore

1.1 Abstract:
In this study, DSBot—a user-friendly assistant designed for those who aren't professionals in data science—is introduced. Through the use of conversational interfaces and AutoML, DSBot surpasses TPOT in 19 out of 30 datasets and attains a BLEU score of 0.8 in translation tasks. Users can formulate research questions, convert them into operational pipelines, and carry out independent end-to-end data analysis with it.

1.2 Introduction:
DSBot builds on GeCoAgent's success by addressing the lack of competence in Data Science. It makes use of conversational technologies, NLP, and AutoML to allow research questions to be declaratively expressed in a domain-independent manner. Empirical testing reveals that DSBot performs comparably to TPOT with much shorter execution times, highlighting its inventiveness in improving efficiency and accessibility for non-experts.

1.3 State of the Art:
The article covers developments in AutoML, Interactive Machine Learning, and Automatic Code Generation. It emphasizes how data science tools are changing and how important it is to have tools like DSBot that can help users who are not experts or experts communicate with each other.

1.4 Methods:
Using user interactions, a language appropriate to the domain, and automation, DSBot simplifies data analysis chores. The eight-step procedure includes automated modules for categorization and regression, conversational comprehension assessment, question translation, and dataset analysis.

1.5 Use Case:
The chat use case highlights DSBot's domain-agnostic capabilities, while the analytical use case illustrates its applicability in genomic data. Both examples highlight how easily DSBot may lead people through the data analysis process.

1.6 Evaluation:
Comparison of DSBot's performance against TPOT shows that for both classification and regression tasks, it performs better in terms of accuracy and computation time. An efficient translation from English to DAW is shown by the GPT-2-based translator, which receives a BLEU score of 0.8.

1.7 Conclusion:
With the help of DSBot, anyone can perform intricate analyses on their own, making data science more accessible to a wider audience. The paper presents potential and empirical evidence, but it also identifies limitations, including imbalanced datasets, and plans for future improvements, such as better conversational features and support for more algorithms. A big step in the direction of greater accessibility in data science is represented by DSBot.

2.1 Constraints:
Although DSBot exhibits great potential, it is limited by possible biases in machine learning models, particularly when addressing imbalanced datasets. Expanding supported pipelines and operations and improving machine learning model selection criteria may also be beneficial to the system.

3.1 Synthesis:
DSBot proves to be a potent and user-friendly instrument that democratizes data science by empowering non-specialists to carry out intricate analyses on their own. Its potential impact is demonstrated by its success in comparative evaluations and user-friendly interfaces. Still, admitting shortcomings draws attention to how difficult it is to maintain fairness and improve functionality. The paper presents DSBot as a developing solution for wider data science accessibility, with a focus on future developments, such as enhanced conversational capabilities and support for additional algorithms.