

## Generating Insights from Data Using a Hybrid Approach

**Authors :** Allmin Susaiyah, Aki Härmä, Milan Petković

**Abstract :** Automatic generation of insights from data using insight mining systems (IMS) is useful in many applications, such as personal health tracking, patient monitoring, and business process management. Existing IMS face challenges in controlling insight extraction, scaling to large databases, and generalising to unseen domains. In this work, we propose a hybrid approach consisting of rule-based and neural components for generating insights from data while overcoming the aforementioned challenges. Firstly, a rule-based data 2CNL component is used to extract statistically significant insights from data and represent them in a controlled natural language (CNL). Secondly, a BERTSum-based CNL2NL component is used to convert these CNLs into natural language texts. We improve the model using task-specific and domain-specific fine-tuning. Our approach has been evaluated using statistical techniques and standard evaluation metrics. We overcame the aforementioned challenges and observed significant improvement with domain-specific fine-tuning.

**Keywords :** data mining, insight mining, natural language generation, pre-trained language models

**Conference Title :** ICDMBDDDE 2023 : International Conference on Data Mining, Big Data, Database and Data Engineering

**Conference Location :** Amsterdam, Netherlands

**Conference Dates :** May 04-05, 2023