Information Extraction on Crypto Project Whitepapers to Assess Credibility and Support Investor Decision-Making

Ery Setiyawan Jullev Atmadji, S.Kom, M.Cs *as a supervisor*

Arya Fattah Duta Sanjaya
Study Program of Informatics Engineering
Majoring of Information Technology

ABSTRACT

A whitepaper is the primary document used by cryptocurrency projects to introduce their vision, technology, and economic model to potential investors. Given the high degree of information asymmetry and the frequent occurrence of misleading or plagiarized content in whitepapers, investors require objective tools to assess project credibility. This study aims to develop a system based on Natural Language Processing (NLP) that can automatically perform information gathering and keyword extraction from whitepaper content. The analysis process includes text extraction from PDF files, text preprocessing, content classification using transformer models (BERT), detection of linguistic errors, and keyword extraction using TF-IDF method. The analysis results are presented through visualizations such as word clouds and highlighted key information, along with assessment based on five data-based signals to provide an initial credibility assessment. Through this approach, the system is expected to support investment decision-making in a more informative, efficient, and data-driven manner. This research demonstrates that the integration of NLP in whitepaper analysis has the potential to enhance transparency and accountability within the cryptocurrency ecosystem.

Key words: *Whitepaper*, NLP, *Cryptocurrency*, BERT, *Keyword Extraction*.