World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:18, No:12, 2024

Early Impact Prediction and Key Factors Study of Artificial Intelligence Patents: A Method Based on LightGBM and Interpretable Machine Learning

Authors : Xingyu Gao, Qiang Wu

Abstract : Patents play a crucial role in protecting innovation and intellectual property. Early prediction of the impact of artificial intelligence (AI) patents helps researchers and companies allocate resources and make better decisions. Understanding the key factors that influence patent impact can assist researchers in gaining a better understanding of the evolution of AI technology and innovation trends. Therefore, identifying highly impactful patents early and providing support for them holds immeasurable value in accelerating technological progress, reducing research and development costs, and mitigating market positioning risks. Despite the extensive research on AI patents, accurately predicting their early impact remains a challenge. Traditional methods often consider only single factors or simple combinations, failing to comprehensively and accurately reflect the actual impact of patents. This paper utilized the artificial intelligence patent database from the United States Patent and Trademark Office and the Len.org patent retrieval platform to obtain specific information on 35,708 AI patents. Using six machine learning models, namely Multiple Linear Regression, Random Forest Regression, XGBoost Regression, LightGBM Regression, Support Vector Machine Regression, and K-Nearest Neighbors Regression, and using early indicators of patents as features, the paper comprehensively predicted the impact of patents from three aspects: technical, social, and economic. These aspects include the technical leadership of patents, the number of citations they receive, and their shared value. The SHAP (Shapley Additive exPlanations) metric was used to explain the predictions of the best model, quantifying the contribution of each feature to the model's predictions. The experimental results on the AI patent dataset indicate that, for all three target variables, LightGBM regression shows the best predictive performance. Specifically, patent novelty has the greatest impact on predicting the technical impact of patents and has a positive effect. Additionally, the number of owners, the number of backward citations, and the number of independent claims are all crucial and have a positive influence on predicting technical impact. In predicting the social impact of patents, the number of applicants is considered the most critical input variable, but it has a negative impact on social impact. At the same time, the number of independent claims, the number of owners, and the number of backward citations are also important predictive factors, and they have a positive effect on social impact. For predicting the economic impact of patents, the number of independent claims is considered the most important factor and has a positive impact on economic impact. The number of owners, the number of sibling countries or regions, and the size of the extended patent family also have a positive influence on economic impact. The study primarily relies on data from the United States Patent and Trademark Office for artificial intelligence patents. Future research could consider more comprehensive data sources, including artificial intelligence patent data, from a global perspective. While the study takes into account various factors, there may still be other important features not considered. In the future, factors such as patent implementation and market applications may be considered as they could have an impact on the influence of patents.

Keywords : patent influence, interpretable machine learning, predictive models, SHAP
Conference Title : ICDS 2024 : International Conference on Data Science
Conference Location : Sydney, Australia
Conference Dates : December 02-03, 2024