RV College of Engineering®

Department of Artificial Intelligence and Machine Learning

Big Data Technologies(AI362IA)

TITLE	Automated Underwriting Engine using Big Data and ML
TEAM	1RV22AI051-Sharankrishna Kondi 1RV22AI053-Shivukumar Hotagi 1RV22AI043-Sandeep S Pawar

INTRODUCTION

As of 2024, the Indian insurance industry has grown into a Rs. 10.8 lakh crore market, with the life insurance sector contributing over Rs. 7.8 lakh crore in premium income and the general insurance sector generating more than Rs. 3 lakh crore, according to the IRDAI. Despite this growth, insurance penetration remains low at 4.2% of GDP, compared to the global average of 7.4%. Moreover, insurance density in India is only USD 91, far below the world average of USD 391, indicating significant room for expansion. The number of lives covered under health insurance crossed 58 crore in FY 2023, showing growing demand for coverage but also emphasizing the need for efficient processing to handle increasing volume. With the exponential rise in policy applications and claims, insurers are facing massive data inflows. The industry now handles petabytes of structured and unstructured data, including medical histories, transaction logs, claim records, social media behavior, and telematics. A 2023 Deloitte India report highlighted that over 70% of Indian insurers are prioritizing Big Data analytics to manage risk, reduce fraud, and improve customer targeting. Furthermore, studies show that integrating data-driven decision-making into underwriting can lead to a 20–30% improvement in risk accuracy, helping insurers offer more personalized and competitive products. Machine Learning adoption in underwriting is accelerating across the Indian insurance ecosystem. According to a 2024 NASSCOM and BCG survey, 63% of insurers in India have implemented or are piloting AI/ML models in their underwriting workflows. These models have shown to reduce underwriting time by 50–70%, and lower manual error rates by up to 25%, resulting in faster policy issuance and improved customer satisfaction. Additionally, insurers using automated risk engines have reported a 15–20% reduction in fraudulent claims. These trends strongly support the development of intelligent underwriting systems that can scale with India's growing demand for inclusive and efficient insurance services

EXISTING SYSTEMS

The existing manual underwriting system in Indian insurance operates through a traditional paper-based framework where human underwriters process applications sequentially using standardized forms and limited data analysis. Current processes involve multiple stakeholders including agents, underwriters, and medical examiners who manually evaluate 20-30 variables per application such as age, income, and medical history using legacy IT infrastructure, Excel spreadsheets, and basic CRM tools. Individual underwriters handle only 75-120 applications daily with processing times ranging from 7-30 days, relying on static actuarial tables and historical risk assessment matrices that lack real-time market adaptability. The workflow follows linear progression through screening, document collection, risk classification, and approval stages with extensive quality control reviews that further extend operational timelines.

PROPOSED SYSTEM

The proposed system is an Automated Underwriting Engine that leverages Big Data technologies like Apache Hadoop and Spark alongside Machine Learning algorithms to streamline and optimize the insurance underwriting process. It is designed to ingest and process large volumes of heterogeneous data—including applicant demographics, health metrics, claim histories, and risk indicators—to build predictive models that assess risk scores in real-time. The system will automate decision-making by classifying applicants into risk categories, identifying potential fraud, and recommending policy approval or rejection instantly. By simulating realistic insurance scenarios with publicly available datasets, the engine demonstrates how intelligent automation can enhance underwriting efficiency, reduce manual intervention, and enable faster, data-driven insurance issuance, especially crucial for a digitally expanding Indian market.

SOCIETAL CONCERNS

While automated underwriting brings speed and efficiency, it raises important societal concerns around data privacy, algorithmic bias, and transparency. In the Indian context, where data protection laws are still evolving, the misuse or leakage of sensitive personal and health data could lead to ethical and legal challenges. Moreover, if ML models are trained on biased or incomplete data, they may unintentionally discriminate against certain groups, leading to unfair denial of coverage. Ensuring ethical AI practices, transparency in decision-making, and strict adherence to data privacy norms is essential to build public trust and ensure equitable access to insurance for all segments of society.