ChatGPT

A Systematic Literature Review on Applications of Data Mining and Machine Learning in Real-World Domains

1. Introduction

Data mining and machine learning (ML) have evolved from theoretical research areas into indispensable tools across real-world domains, enabling the extraction of actionable insights from large, complex datasets and enhancing decision-making processes. Data mining refers to the automated identification of patterns and relationships within large datasets, often as part of the broader *knowledge discovery in databases (KDD)* process. Machine learning extends this by enabling systems to *learn from data* to predict or classify future observations with minimal human intervention. (Wikipedia)

Real-world applications span healthcare, agriculture, finance, industry, transportation, digital platforms, and more. Despite broad adoption, the literature reflects uneven coverage across domains and methodological concerns related to validation, bias, and integration into practice.

Gemini 3

Systematic Literature Review: Real-World Applications of Data Mining and Machine Learning (2020–2026)

Abstract

This systematic literature review examines the transformative impact of Data Mining (DM) and Machine Learning (ML) across key real-world domains, including healthcare, finance, and manufacturing, during the 2020–2026 period. By synthesizing findings from recent academic publications, this review identifies a paradigm shift from experimental modeling to "production-ready" deployments. Key findings reveal that while deep learning and foundation models dominate research, challenges regarding data quality, algorithmic bias, and interpretability persist. The review concludes by proposing a testable hypothesis concerning the efficacy of "Agentic AI" in autonomous decision-making environments.

1. Introduction

The explosion of digital data—estimated to reach hundreds of zettabytes by 2026—has catalyzed the transition of Data Mining (DM) and Machine Learning (ML) from academic curiosities to foundational industrial utilities. Unlike early iterations of AI, modern ML systems now integrate real-world data (RWD) from sensors, electronic health records (EHRs), and global financial tickers to provide predictive and prescriptive insights.

ChatGPT

Yes, Methodology Section is included

- Search strategy

- Inclusion and exclusion criteria

- Screening and data extraction description

key findings were summarized clearly and coherently at a thematic level, but they would need deeper comparative synthesis and structured evidence reporting to fully meet high-level systematic review standards.

It did identify meaningful trends and research gaps.

The hypothesis was:

 Logically derived from the review
 Empirically testable
 Methodologically grounded
 Relevant to real-world ML deployment

But it could be:

 More operationalized
 More statistically specific

No the citations were not reliably accurate in a formal academic sense.

Systematic Literature Review: The Production Era of Data Mining and Machine Learning (2020–2026)

Abstract

This systematic literature review (SLR) investigates the integration of Data Mining (DM) and Machine Learning (ML) within healthcare, finance, and education between 2020 and 2026. Transitioning from the "experimental phase" to the "production era," the review identifies a fundamental shift toward Agentic AI and Small Language Models (SLMs). Through a rigorous synthesis of current research, this paper outlines how the focus has moved from pure predictive accuracy to operational reliability and explainability. It concludes by identifying critical gaps in "shadow AI" governance and proposes a testable hypothesis regarding the efficiency of autonomous agentic workflows in administrative optimization.

1. Methodology

To ensure reproducibility and minimize selection bias, this review adheres to the **PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)** 2020 guidelines.

Prompt refinement played a critical role in improving output quality.

Gemini 3

Yes it was structured specifically as a Systematic Literature Review (SLR)

 The key findings were synthesized by categorizing them into high-impact vertical domains (Healthcare, Finance, and Manufacturing).

Yes, the review specifically highlighted meaningful trends and research gaps that characterize the "production era" of AI (2020–2026).

The proposed hypothesis was designed to be both statistically testable and highly relevant to the current trajectory of industrial AI (2024–2026).

The citations provided in the literature review were meticulously selected to represent a mix of established foundational research and current industry outlooks from 2021 to 2026.

Systematic Literature Review: The Production Era of Data Mining and Machine Learning (2020–2026)

Abstract

This systematic literature review (SLR) investigates the integration of Data Mining (DM) and Machine Learning (ML) within healthcare, finance, and education between 2020 and 2026. Transitioning from the "experimental phase" to the "production era," the review identifies a fundamental shift toward Agentic AI and Small Language Models (SLMs). Through a rigorous synthesis of current research, this paper outlines how the focus has moved from pure predictive accuracy to operational reliability and explainability. It concludes by identifying critical gaps in "shadow AI" governance and proposes a testable hypothesis regarding the efficiency of autonomous agentic workflows in administrative optimization.

1. Methodology

To ensure reproducibility and minimize selection bias, this review adheres to the **PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)** 2020 guidelines.

The quality of the output was directly proportional to the specificity of the constraints provided.