iJRASET

International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: IV Month of publication: April 2025

DOI: https://doi.org/10.22214/ijraset.2025.68704

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

Machine Learning: The Cognitive Refinery for Big Data

Sanjana[1], Paras Verma[2]
[1, 2]Mechanical Engineering Graduate | Independent Researcher in Machine Learning & Big Data Analytics

Abstract: The rise of big data has significantly reshaped various industries by enabling the analysis of vast and complex datasets that were previously unimaginable. Often characterized by the three Vs—Volume, Velocity, and Variety—these data streams surpass the capabilities of traditional analytical methods. As a result, machine learning (ML) has become an essential tool for extracting actionable insights from such high-dimensional information. This review explores a broad spectrum of ML approaches suitable for big data analytics, examining their advantages, challenges, and real-world applications. The discussion spans supervised, unsupervised, semi-supervised, reinforcement, and deep learning techniques.

It discusses each approach, and their relative applicability to various big data problems, including issues with real-time processing, high-dimensionality, noise tolerance, and computability scalability. The paper also examines case studies and applications of ML in various industries, demonstrating how it can be used to inform decision-making, foster innovation, and create value in sectors such as healthcare, finance, retail, and cybersecurity.

Outside of the present scenario, the focus shifts towards addressing some of the major difficulties in applying ML to large-scale data, including problems with data quality, interpretability of models, resource needs, and ethical dilemmas like bias and privacy. It also discusses emerging trends and future directions that include federated learning, automated machine learning (AutoML) and the merger of ML and edge computing. The objective of the paper is to provide such insights to researchers, practitioners, and decision-makers, from which they can benefit as they endeavor to utilize machine learning to capitalize on the advantage of big data.

Keywords: Deep Neural Architectures, Supervised Algorithms, Semi-Supervised Approaches, Reinforcement-Based Training, Unsupervised Methods, Big Data, Intelligent Systems, Data Science, Hadoop, Apache Spark, Scalability, Decentralized Computing.

I. INTRODUCTION

The rapid expansion of technologies like IoT devices, social networks, cloud infrastructure, and mobile computing has significantly accelerated data generation across industries. As a result, we have entered an era where traditional data handling methods are no longer sufficient to manage the vast and complex datasets being produced. This evolution is often framed in terms of the five Vs of big data: Volume (data quantity), Velocity (speed of creation), Variety (types and sources), Veracity (reliability and quality), and Value (practical usefulness). While storing data is fundamental, it is only the beginning—advanced analytics are now essential for extracting meaning.

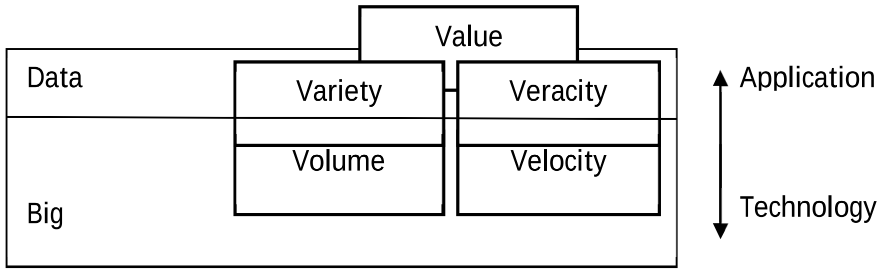
To make sense of this data deluge, machine learning (ML) plays a critical role. ML, a subset of artificial intelligence, enables the discovery of patterns, trends, and insights by automatically learning from large-scale data [1]. This makes it especially effective in environments that demand high-speed, high-volume processing.

II. LITERATURE REVIEW

In recent years, we've entered what is commonly referred to as the "Big Data era," where the size and complexity of data are growing at an unprecedented pace.

This shift has brought both opportunities and challenges, especially when it comes to making informed decisions based on data insights. Key characteristics that define big data are often categorized as the **five Vs**: **volume** (quantity), **velocity** (speed), **variety** (types), **veracity** (reliability), and **value** (usefulness). Traditional data processing systems are not equipped to handle this level of scale. In contrast, machine learning (ML) has emerged as a powerful approach for automatically detecting patterns and generating predictive models—without requiring manual programming [1].

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com



A. Why Machine Learning is Essential for Big Data

A number of researchers note that the real worth of Big Data can be achieved only by means of intelligent analysis, made possible by ML. In the words of Chen, Mao, & Liu (2014) [2], ML brings with it the means by which we can expose hidden patterns and associations in vast, diverse data sets. The automation ML brings also cuts down human labor enormously, while enhancing the scalability, a dimension vital while working with data sets by petabytes or exabytes.

ML also facilitates adaptive learning—models can be constantly modified as more data become available. This is especially useful in highly dynamic contexts such as e-commerce, finance, and medicine, where conditions constantly change. Furthermore, most ML algorithms can deal with noisy or missing data, solving one of the major drawbacks of traditional statistical approaches when used in real-life Big Data contexts [3].

B. How Machine Learning Improves Big Data Handling

Machine learning contributes to the effective handling of big data in several impactful ways:

1) Data Preparation and Feature Optimization

Cleaning datasets and extracting relevant features are fundamental steps in data analysis. ML techniques have proven useful in automating and refining these stages. Algorithms such as decision trees, clustering models, and neural networks help isolate key attributes, minimize noise, and improve dataset reliability [4]. For instance, in unsupervised learning, approaches like k-means and hierarchical clustering are commonly used to group similar data instances, which supports both segmentation and anomaly detection tasks.

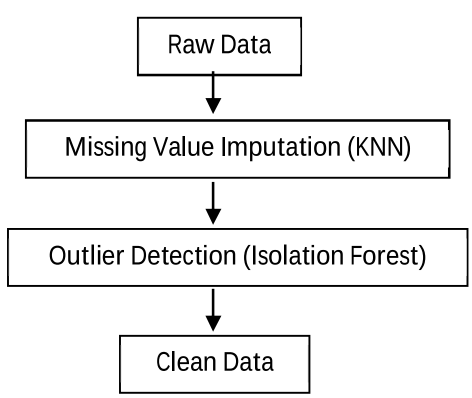
Major Applications:

- Dirty Data Cleaning: Techniques in ML, for instance, the K-Nearest Neighbors (KNN), can impute missing values in a smart way, while the likes of Isolation Forest can eliminate outliers.
- Feature Engineering: ML recognizes and builds informative features that most accurately reflect underlying data patterns.
- Dimensionality Reduction**:** When working with high-dimensional datasets, techniques like Principal Component Analysis (PCA) and t-SNE are applied to simplify data structure for easier visualization and interpretation.

Case Study: Google Cloud Data preparation

Google's DataPrep tool applies ML to prepare petabyte-sized data stored in BigQuery [8]. The tool performs repetitive tasks, learns from user behavior, and cuts down human labor by more than 70%.

Flowchart: ML-Based Data Cleaning Pipeline



International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

2) Predictive Modeling and Forecasting

ML is adept at future forecasting by learning from past data. Support vector machines, decision trees, regression models, and ensemble-based approaches such as random forests are commonly employed in Big Data contexts for forecasting trends, customer sentiment, or failures in a system. Models such as LSTMs (Long Short-Term Memory networks) are used in time-series data, yielding a high degree of accuracy in forecasting temporal patterns [5].

Supervised Learning (Classification & Regression):

- Example1: PayPal Fraud Detection

Employs Random Forest and XGBoost to process millions of transactions within a year, detecting more than $6 billion in fraud.

- Example 2: Demand Forecasting (Walmart)

Deploys LSTM Neural Networks with a predictive accuracy of up to 95% for inventory demand, facilitating inventory and logistics optimization.

Unsupervised Learning (Clustering & Anomaly Detection)

- Example 1: Netflix User Segmentation

K-Means Clustering is employed by Netflix in segmenting users according to behavior and preferences. This enhances user-specific recommendations as well as targeted content

3) Real-time Analytics and Decision Support

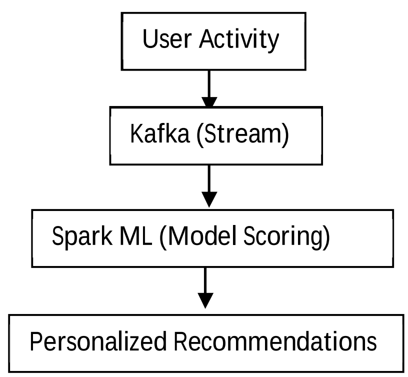
One of the greatest value additions from combining ML with Big Data is in real-time decision-making. ML can process real-time streaming data, and this allows companies to respond in real time when things occur.

Examples are recommendation systems based on users' behavior (e.g., Amazon, Netflix), as well as fraud detection systems identifying anomalies in real time [6].

Recommendation Systems

- Amazon uses collaborative filtering in combination with deep learning in product recommendations—accounting for 35% of overall sales.
- Spotify applies NLP and reinforcement learning to create playlists based on listener context and mood.

Schematic: Real-Time Recommendation Engine



Fraud Detection in Banking

- Deep learning-based systems integrated into payment gateways are used by JPMorgan Chase to stop more than 100,000 fraudulent payments each day, safeguarding users in real-time.

4) Personalization and Customer Insights

ML algorithms examine large volumes of user data in order to personalize content, offerings, and services according to user desires. Collaborative filtering, clustering, and NLP methods allow sites to deliver customized experiences, resulting in user engagement and conversion rates. This has worked best in business fields such as digital advertising, e-commerce, and entertainment (Ricci et al., 2011).

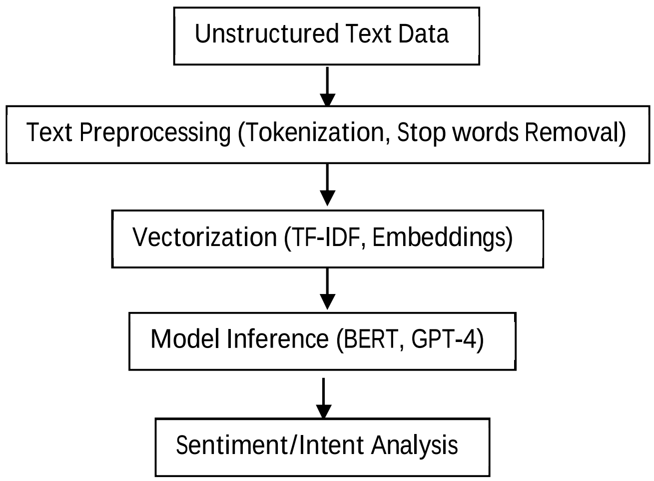
- Twitter sentiment analysis

ML models like BERT and GPT-4 process over 500 million tweets daily, identifying public opinion trends and brand sentiment.

- Google Search (RankBrain)

Leverages ML to handle more than 8.5 billion daily searches, interprets context, and enhances relevance in results.

Schematic: Real-Time Recommendation Engine



Summary

The literature repeatedly emphasizes how Machine Learning is a key driver in extracting value from Big Data. It supplies the automation, scalability, responsiveness, and smarts required to convert raw, unstructured data into useful knowledge. Big Data supplies the inputs—the "fuel"—while ML is the "engine" that propels analysis, prediction, and decision-making.

With the continuing expansion in data volume and sophistication, the integration of ML in Big Data ecosystems is likely to become even more critical. Future studies can be directed towards enhancing interpretability of the model, addressing bias issues, as well as generating hybrid models where a combination of ML approaches are employed in addressing emerging data issues more effectively.

III. MACHINE LEARNING IN BIG DATA ANALYTICS

A. Supervised Learning

Supervised learning involves training a model using **labeled datasets**, where input-output pairs are known, to enable the system to make predictions on new, unseen data. It is particularly effective for both classification and regression tasks in big data scenarios, offering high levels of accuracy and scalability.

Methods and Tools

- Linear and Logistic Regression: Efficient for big data classification, particularly with scalable tools such as Apache Spark MLlib [7]. Widely employed in customer churn prediction, price models, credit scoring.
- Support Vector Machines (SVM): High-dimensional data, usually supported by the use of Hadoop, together with MapReduce frameworks for dealing with computations.
- Decision Trees & Random Forests: Offer interpretability and scalability. Decision tree ensembles are popular in fraud detection, as well as customer segmentation.

B. Unsupervised Learning

Unsupervised learning extracts information from data with no explicit labels, thus being specifically designed for exploratory analysis in large, unstructured datasets.

Methods and Tools

- K-Means Clustering: Commonly employed for customer segmentation; can be scaled with Spark MLlib K-Means.
- Hierarchical Clustering: While being more computationally demanding, it yields in-depth insights in nested data structures.

- Principal Component Analysis (PCA): Reduces dimensionality of data, enhancing processing efficiency while maintaining key information.

C. Semi-Supervised Learning

Such a strategy relies both on unlabeled data as well as labeled data, most advantageously in big data contexts where the latter is costly or scarce.

Methods and Tools

- Co-Training and Self-Training: Iteratively label new, unlabelled data, upping training speed. Can scale with batch-processing software such as Hadoop.
- Graph-Based Models: Beneficial in social network analysis, leveraging tools such as GraphX (Spark) for relational data modeling.

D. Reinforcement Learning (RL)

RL methods learn from their environment by interacting and getting feedback. The adaptive feature is useful in dynamic big data systems.

Methods and Tools

- Q-Learning and Deep Q-Networks (DQN): Used in recommendation systems and real-time auctions in online advertising.
- Policy Gradient Methods: Applied in robotics, autonomous vehicles, and finance trading; scalable through multi-agent distributed systems.

E. Deep Learning (DL)

Deep learning algorithms derive high-level features from raw data. Deep learning has expanded exponentially with the arrival of GPUs and cloud computing.

Methods and Tools

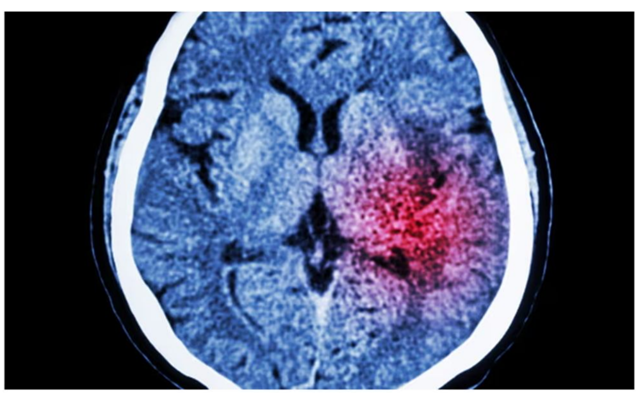
- Convolutional Neural Networks (CNNs): Strong in facial recognition, video analytics, and image classification.
- Recurrent Neural Networks (RNNs) & LSTMs: Best suited for time series forecasting and natural language processing (NLP), employed in sentiment analysis, chatbots, and predictive maintenance.
- Transformer Models: These include architectures such as BERT and GPT, whose performance is boosted by multi-GPU and distributed computer configurations.

IV. CASE STUDY

A. Improving Stroke Prediction in Healthcare

Background: Atrial fibrillation, or AF, is a condition of the heart that hugely raises the incidence of stroke. In the UK, there are around 1.6 million people with AF, with many more perhaps not yet diagnosed, resulting in avoidable strokes.

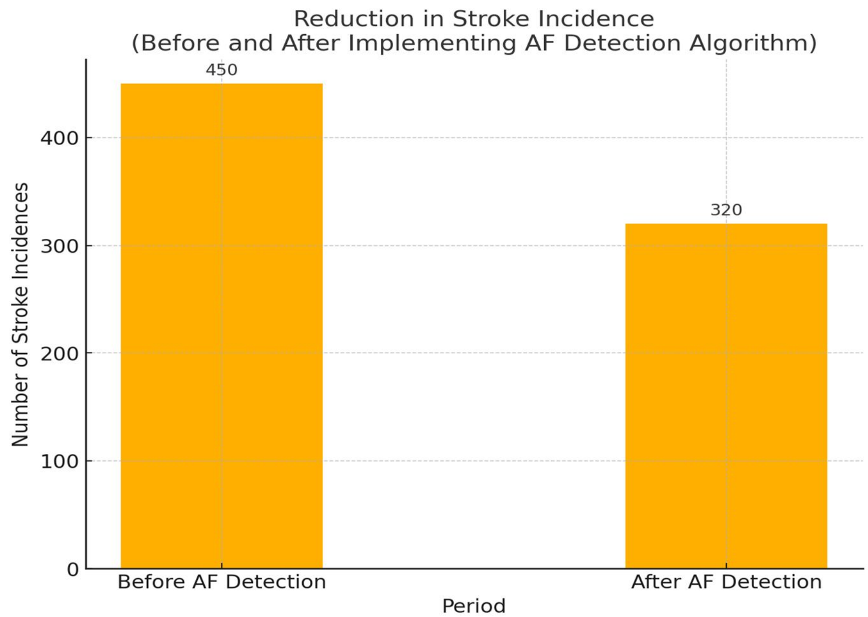
Solution: A research team from a major UK healthcare trust, working in partnership with a leading academic institution, developed a machine learning model to assess the likelihood of atrial fibrillation (AF) in patients. Using anonymized medical records collected from general practitioners, the model evaluates various risk factors such as age, gender, ethnicity, and medical history to generate predictions.



International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

Implementation: The NHS initiated the Find-AF project, backed by key healthcare organizations, to trial this predictive tool across regions in West Yorkshire [9]. Patients identified as high-risk by the algorithm receive ECG monitoring for atrial fibrillation, enabling timely medical intervention.

Impact: Potential early detection by this machine learning solution can stop thousands of strokes per year, enhance outcomes for patients, and lower costs for the health system.



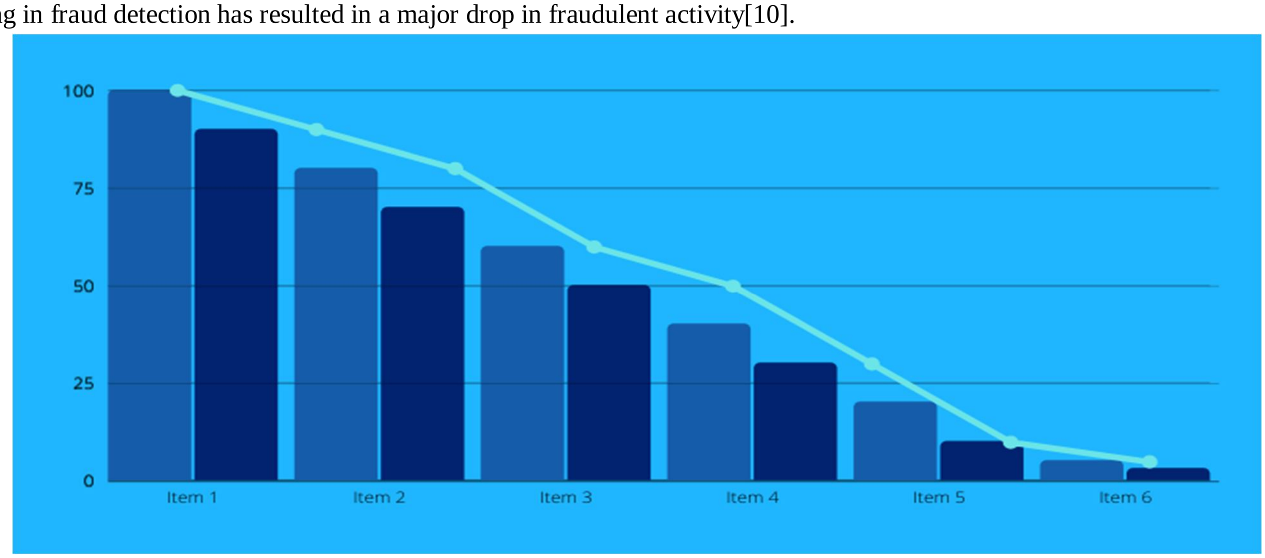
B. Fraud Detection in Financial Services

Background: Financial institutions are confronted with serious challenges in detecting fraud and preventing its occurrence, resulting in immense monetary losses, as well as reputational harm.

Solution: Banks use machine learning tools to track consumer spending habits in real-time. The tools use algorithms to detect suspicious behavior, potentially signaling fraud.

Implementation: Machine learning algorithms can identify anomalies in large sets of transactional records, for instance, uncharacteristically large purchases or payments made in uncommon locations. Once identified, the system can flag or freeze the account, eliciting follow-up investigations.

Impact: This proactive strategy allows banks to operate quickly, saving losses for the bank as well as clients. Use of machine learning in fraud detection has resulted in a major drop in fraudulent activity[10].



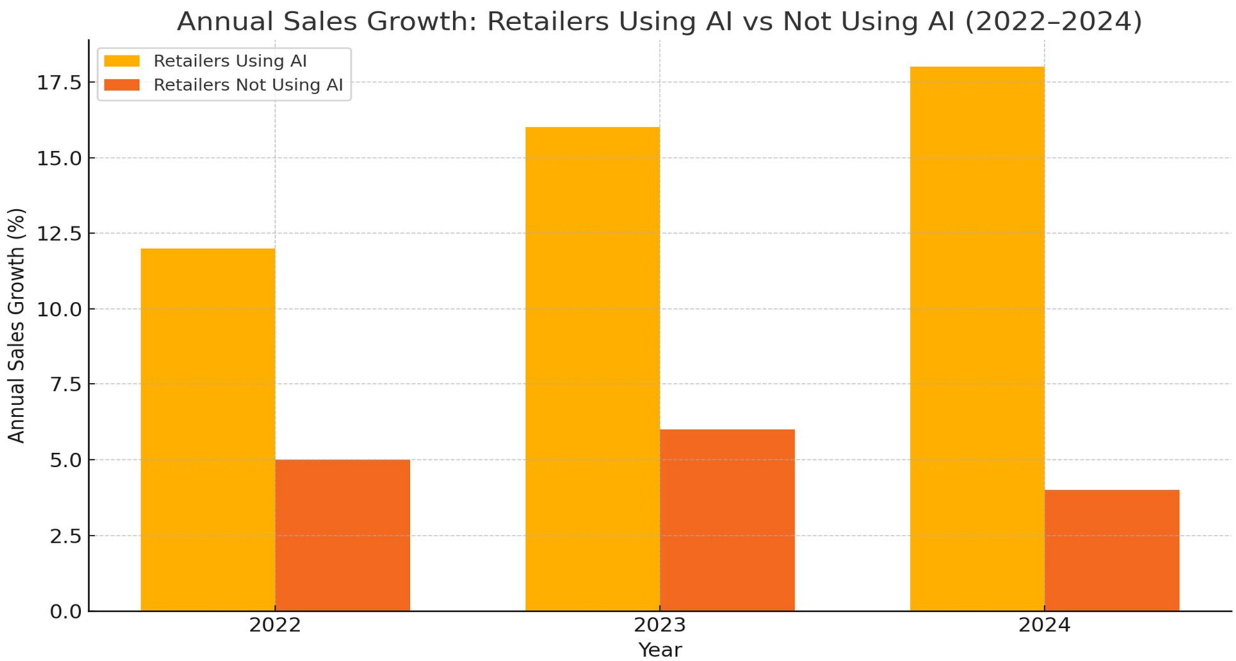
C. Personalization in Retail Marketing

Retailers are looking to maximize customer engagement, as well as increase sales, with tailor-made marketing strategies. The traditional methods tend to struggle in being able to handle large volumes of customer data.

The solution: Retail companies leverage machine learning techniques to analyze consumer data—such as previous purchases, online activity patterns, and demographic profiles—to deliver personalized marketing messages and product recommendations.

Implementation: Consumer segmentation can be more precisely executed by retailers through the use of big data analytics, with offerings customized according to individual likings. For example, predictive modeling by machine learning can identify the items a customer would probably be interested in and offer them through personalized emails or app push notifications.

Impact: The retailers who are leveraging AI and machine learning technology have seen remarkable performance improvements. From 2022 to 2024, firms that use these technologies achieved double- digit growth in sales over the past years, surpassing competitors not adopting AI solutions.



These examples illustrate the power of machine learning tools in big data in transforming industries across the board, resulting in better results, higher efficiency, and substantial costs savings.

V. CHALLENGES & SOLUTIONS IN BIG DATA: ROLE OF MACHINE LEARNING

While integrating Big Data with Machine Learning (ML) has led to remarkable breakthroughs, it also introduces significant challenges. Below is a breakdown of key issues and the ML-driven strategies used to overcome them.

Challenge	ML Solution	Real-World Example
Data Volume Exceeds Limit	Distributed Machine Learning	Uber's Michelangelo Platform
Real-time Processing Requirements	Online Learning Algorithms	Twitter's Real-Time Trends System
Unstructured Data	Deep Learning (CNNs, RNNs, Transformers)	Facebook’s Image & Video Recognition

Problem: Data Volume Too Great

With organizations gathering petabytes of data by the day, training ML models one machine at a time is no longer possible due to memory, time, and compute limitations.

A. Challenge: Real-Time Processing Requirements
Many applications—such as fraud detection, personalized recommendations, and social media analysis—require responses with minimal delay, often in real or near-instant timeframes. Traditional batch-processing models are too slow to handle such demands effectively.

B. ML Solution
1) Online Learning (Stochastic Gradient Descent)
Online learning also has the benefit of enabling continuous model update as new data is received, instead of training from scratch. Stochastic Gradient Descent (SGD) along with its variants facilitate rapid model update and low-latency response.
Example: Twitter's Real-Time Trends System
Twitter applies online learning models for:
• Identifying top trending hashtags and topics worldwide in a matter of seconds.
• Updating user interest models constantly through new tweets and user interactions.
• Achieving scalability and speed even with more than 500 million tweets/day.

2) Distributed Machine Learning
Distributed ML frameworks, such as TensorFlow on Apache Spark, split data and computations across a cluster's many nodes. This parallelizes training of the model and allows for processing of vast data sets.
Example: Uber’s Michelangelo Platform
Uber built Michelangelo, a scalable ML-as-a-service platform, which:
• Enables distributed training of TensorFlow models on Spark.
• Supports use cases such as ETA prediction, fraud detection, and dining recommendations.
• Processes billions of predictions daily within Uber's worldwide network.

C. Unstructured Data Handling Challenge
More than 80% of Big Data is unstructured, such as images, video, text, and sound. This data necessitates more than the use of traditional ML methods.
Deep learning techniques are particularly effective at processing unstructured information:
- Convolutional Neural Networks (CNNs): Commonly applied to visual data such as images and videos.
- Recurrent Neural Networks (RNNs) and Transformers: Used extensively for analyzing textual inputs and patterns over time.
- Autoencoders and GANs: Utilized in synthetic data creation and reducing dimensional complexity.

Example: Facebook’s Image Recognition System
Facebook applies CNN-based models to:
• Auto-tag people and items in uploaded photos.
• Provide advanced facial recognition and content filtering functionalities.
• Process billions of images and videos every day across its platforms.

VI. FUTURE TRENDS IN MACHINE LEARNING FOR BIG DATA
With industries reaching maturity in the adoption of ML and Big Data, the future wave of innovation is likely to see a drive for scalability, automation, privacy, and data democratization. The upcoming, emerging trends mirror this evolution and are likely to change the way data-driven systems behave in the future.
A. Federated Learning: Privacy-Preserving Intelligence at the Edge
Federated learning (FL) is a new training paradigm for ML methods, where data is not centralized but the model is trained locally on devices (e.g., smartphones) and the model parameter updates are shared with a central server. This preserves data privacy while also saving communications costs.
Main Benefits:
• Safeguards user privacy (raw data does not leave device)
• Minimizes latency and bandwidth consumption
• Best for industries such as healthcare, banking, and mobile apps

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

Real-Life Scenario:
Google's Gboard Keyboard employs Federated Learning in order to enhance next-word prediction [11] without sharing users' messages in the cloud.

Hypothetical Forecast (2030)
• 2023: FL in production = 5% of ML deployments
• 2030: Expected to reach 40%, especially in regulated sectors (HIPAA, GDPR)
• Market Size: Estimated at $5.5 billion (from $210M in 2021, CAGR ~47%)

B. AutoML: Automating the ML Pipeline

Automated Machine Learning (AutoML) tools make the best model architecture selection, hyperparameters, and model deployment automatic — essentially eliminating the requirements for ML expertise.
Chief Benefits:
• Democratizes ML for non-technical users
• Shortens model development time from weeks to hours
• Works seamlessly in cloud environments

Real-Life Scenario:
Google Cloud AutoML allows businesses to create high-quality, tailor-made models in a code-free manner.

Projected Assessment (2030):
• 2023: AutoML in ~10% of enterprise AI projects
• 2030: Will power 65% of all enterprise ML use cases
• Time to Deployment: Cut by 80%, enhancing rapid experimentation in sectors like retail, fintech, and logistics

C. AI-Driven Data Lakes: Intelligent, Self-Organizing Storage

Traditional data lakes tend to become "data swamps" without proper governance. The future lies in AI-powered data lakes that tag, index, and handle data automatically using ML.
Major Benefits:
• Improves data discoverability
• Automates governance and lineage tracking
• Enables real-time analytics and supports schema evolution

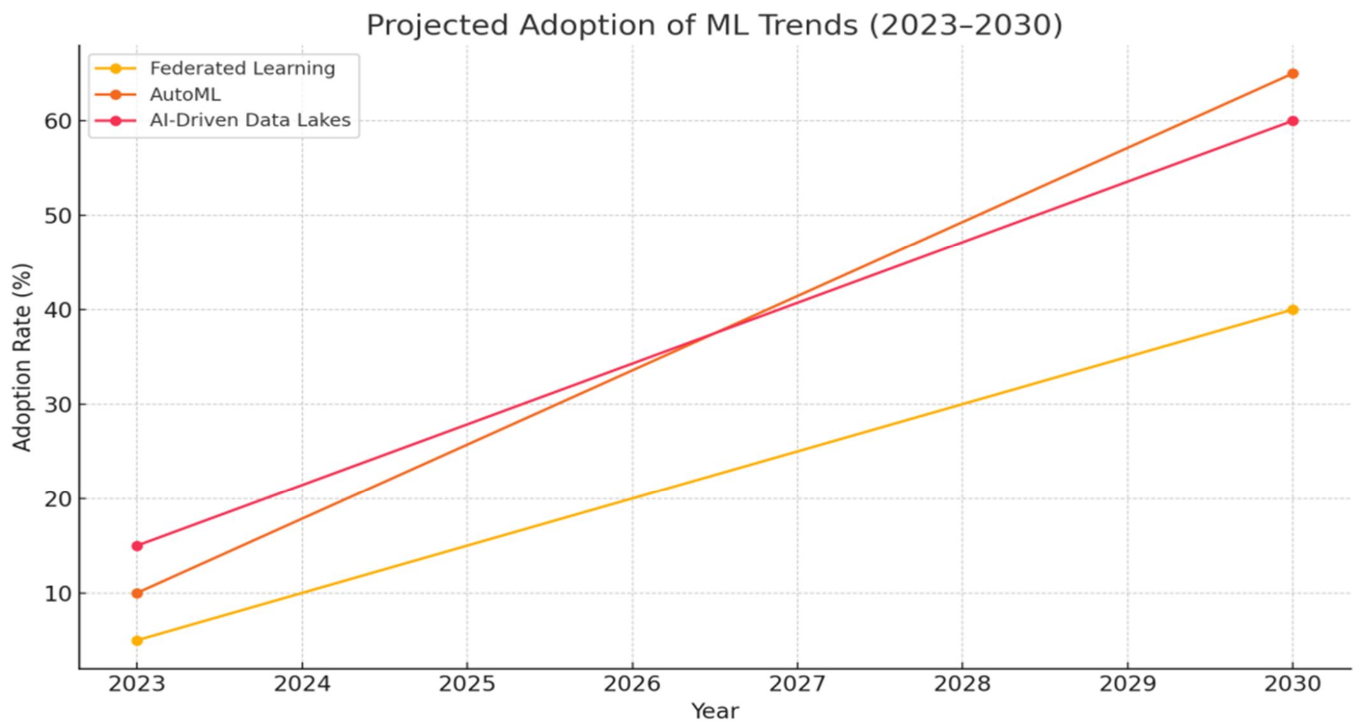
Real-Life Scenario:
Databricks Delta Lake unites ML with cloud data lakes, allowing for real-time query optimization as well as metadata management.

Projected Outlook (2030):
• 2023: Only 15% of data lakes are ML-enabled
• 2030: Over 60% of cloud-native organizations expected to utilize AI-driven data lakes
• Operational Efficiency: Improved by 50–70% due to reduced manual data wrangling

Emerging Trends and Forecasts

Trend	Core Feature	2023 Adoption	2030 Projection	Key Benefit
Federated Learning	Edge-based, privacy-preserving training	5%	40% of ML deployments	Privacy compliance, low-latency ML
AutoML	Automated model development & tuning	10%	65% of enterprise AI use	Democratization of ML, speed to market
AI-Powered Data Lakes	Self-organizing intelligent storage	15%	60% in cloud-native firms	Better governance, real-time access

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com



VII. CONCLUSION: ML AS THE ENGINE OF BIG DATA INTELLIGENCE

Machine Learning is no longer used as a supporting tool—it is the key engine driving Big Data from raw, unstructured data into usable business-ready intelligence fueling decision-making in industry today. The size and sophistication of data keep increasing, further highlighting the role of ML.

A. Scaling up Data Preprocessing

Data cleaning and feature engineering are among the most time-consuming problems in Big Data analytics. These processes are automated by ML algorithms:

• Identification and imputation of missing data through methods such as K-Nearest Neighbors (KNN)
• Detection of outliers by means of Isolation Forests
• Optimal feature selection through Principal Component Analysis (PCA) and Lasso Regression

Impact:

• Cuts manual data preparation time by 60–80% [8]
• Allows analysts and engineers to devote more time to higher-value activities such as model development and business interpretation

B. Predictive Modeling for Forecasting

ML models trained from historical data can generate highly accurate predictions across business functions:

• XGBoost: Widely employed in structured data competitions as well as production pipelines in sales forecasting, credit scoring, and fraud detection
• LSTM Neural Networks: Highly capable for time-series data such as stock prices, supply chain forecasting, energy demand prediction

Real-life examples:

• Walmart: Applies LSTM-based models to predict demand in more than 5,000 locations, increasing inventory availability by 30% [7]
• PayPal: Saves more than $6 billion/year in fraud through XGBoost and outlier detection

C. Real-time decision-making with streaming ML

With the advent of digital platforms and IoT, data comes in real-time. Instant responses are required from systems. ML frameworks in combination with stream processing tools such as Apache Kafka and Apache Spark facilitate this:

• Trained models are used for scoring real-time streaming data
• Firms can take instant decisions—approve a transaction, suggest a product, or trigger a fraud warning

Examples:

• Netflix: Leverages Kafka + Spark + ML to make real-time recommendations responsible for 75% of user engagement
• Visa: Processes more than 65,000 transactions in a second, detecting fraud within 300 milliseconds

Impact on the industry: Synergies

The combination of ML and Big Data is already producing tremendous financial as well as operational impact:

Sector	Case Study	Impact
Healthcare	Disease prediction, medical imaging	Predictive diagnostics lower costs by $150B/year (U.S.)
Finance	Algorithmic trading, credit risk	ML-enabled hedge funds such as Renaissance generate >40% yearly ROI
E-commerce	Recommender systems, dynamic pricing	Up to 35% of sales are driven by personalized experiences (Amazon)
Logistics	Demand forecasting, routing optimization	ML reduces delivery costs by 20–25% (DHL, FedEx)

Closing Thought

"Data is the new oil, and Machine Learning is the refinery that extracts its value."

In today's digital era, as data continues to grow in scale and significance, smart automation stands at the forefront of technological progress. Machine learning enables this data to be transformed—not wasted—turning raw information into meaningful insights that power faster, smarter, and more profitable decisions.

REFERENCES

[1] Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255–260. https://doi.org/10.1126/science.aaa8415
[2] Chen, M., Mao, S., & Liu, Y. (2014). Big Data: A survey. Mobile Networks and Applications, 19(2), 171–209. https://doi.org/10.1007/s11036-013-0489-0
[3] Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, 35(2), 137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007
[4] Zhang, J., Yang, X., & Appelbaum, D. (2015). Toward effective big data analysis in continuous auditing. Accounting Horizons, 29(2), 469–476. https://doi.org/10.2308/acch-51066
[5] Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. IEEE Transactions on Neural Networks, 5(2), 157–166. https://doi.org/10.1109/72.279181
[6] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys, 41(3), 1–58. https://doi.org/10.1145/1541880.1541882
[7] Ricci, F., Rokach, L., & Shapira, B. (2011). Recommender systems handbook. Springer.
[8] Google Cloud. (n.d.). Cloud Dataprep. Retrieved from https://cloud.google.com/dataprep
[9] NHS. (2021). The Find-AF Study. Retrieved from https://www.findaf.org
[10] Uber Engineering. (2017). Introducing Michelangelo: Uber's machine learning platform. Retrieved from https://eng.uber.com/michelangelo-machine-learning-platform
[11] Humby, C. (2006). Data is the new oil. Speech at the ANA Senior Marketers' Conference, Kellogg School of Management.