International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

DATA-DRIVEN INSURANCE CLAIMS ANALYSIS

[1] V Soujanya, [2] P ShivaKeshav, [3] P Abhitha, [4] Y Manoj Kumar
[1]AssistantProfessor, [234]Students
Department of Data Science
Siddhartha Institute of Technology & Sciences, Narapally
vsoujanya.cse@siddhartha.org.in, 23TQ1A6707@siddhartha.co.in, 23TQ1A6728@siddhartha.co.in, 23TQ1A6744@siddhartha.co.in,

Abstract

The Data-Driven Insurance Claims Analysis system is an analytical platform designed to examine insurance claim data and generate meaningful insights from historical records. Insurance companies handle large volumes of claims related to different policies, customers, locations, claim types, and settlement outcomes. Analyzing this information manually can be difficult and time-consuming, making automated data analysis useful for understanding claim patterns and operational performance.

The proposed system collects and processes structured insurance claim information such as claim amount, policy type, claim status, customer details, claim date, location, and settlement information. Data preprocessing techniques are applied to remove duplicate records, handle missing values, correct inconsistent information, and prepare the dataset for analysis. The processed data is then used to calculate important claim-related metrics.

The system provides analytical insights such as total claims, approved and rejected claims, average claim amount, settlement amount, claim frequency, and claim trends over time. It can also analyze claims based on policy categories, geographical regions, customer segments, and claim types. Interactive charts, graphs, tables, and KPI cards make the results easier to understand.

The platform can help identify unusual claim patterns and areas that require further investigation. Trend analysis can highlight changes in claim volume and settlement amounts, while comparative analysis can show differences between policy types or regions. These insights can support insurance teams in monitoring operations, improving reporting, and identifying potential areas for additional review.

Overall, the Data-Driven Insurance Claims Analysis system provides a centralized environment for transforming raw claim records into understandable analytical information. It reduces manual analysis and supports data-driven decision-making. Future enhancements can include predictive claim analysis, anomaly detection, fraud-risk indicators, automated reporting, machine learning models, and real-time integration with insurance management systems.

I. Introduction

The insurance industry manages large amounts of data generated from policies, customers, claims, payments, and settlements. Claims are one of the most important components of insurance operations because they directly affect financial performance and customer service. As the number of policyholders and claims increases, analyzing this information efficiently becomes increasingly important.

Traditional insurance claim analysis often involves spreadsheets, manually prepared reports, and separate operational systems. These approaches can make it difficult to obtain a complete view of claim performance. Manual analysis also requires significant effort when organizations need to compare claims across different periods, policy categories, locations, or customer segments.

Data analytics provides techniques for converting large datasets into useful information. Insurance claim datasets can be analyzed to identify claim frequency,

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

average claim amounts, settlement patterns, rejection rates, and changes over time. Data visualization further improves understanding by presenting important information through charts, graphs, dashboards, and interactive reports.

Data-driven analysis can also support the identification of unusual patterns. For example, a sudden increase in claims within a particular region or policy category may require additional investigation. Analytical systems can highlight such patterns without automatically concluding that a particular claim is fraudulent or invalid. This allows qualified insurance personnel to perform appropriate further review.

The proposed Data-Driven Insurance Claims Analysis system provides a centralized analytical platform for examining insurance claim information. It combines data preprocessing, descriptive analytics, comparative analysis, trend analysis, visualization, and reporting. The system is intended to help users understand historical claim data and support informed operational decisions based on available evidence.

II. Literature Survey

Insurance data analytics has become an important area of research because insurance organizations generate large volumes of structured and semi-structured data. Traditional insurance systems primarily focused on storing policy and claim information for operational purposes. As data volumes increased, organizations began using analytical techniques to understand customer behavior, claims, and financial performance.

Descriptive analytics is commonly used to summarize historical insurance data. Metrics such as claim frequency, average claim amount, settlement amount, approval rate, and rejection rate can provide an overview of insurance operations. Grouping these metrics by policy type, geographical region, claim category, or time period can reveal differences between segments.

Data preprocessing is an important component of insurance analytics. Real-world claim datasets can contain missing values, duplicate records, inconsistent formats, and incorrect data entries. Techniques such as data cleaning, normalization, missing-value treatment, and duplicate removal are therefore used before performing analysis. High-quality preprocessing helps improve the reliability of analytical results.

Data visualization techniques are widely used to communicate insurance insights. Dashboards can display key performance indicators along with charts showing claim trends, settlement distributions, regional patterns, and policy-level comparisons. Interactive visualization allows users to filter information and examine specific segments of the dataset without manually creating multiple reports.

Machine Learning techniques have also been investigated for insurance applications such as claim prediction, risk assessment, customer segmentation, and fraud detection. Classification and anomaly-detection approaches can identify patterns that may require further investigation. However, model outputs should be treated as analytical indicators rather than automatic decisions, particularly when they affect individual customers or claims.

Recent insurance analytics platforms increasingly combine data engineering, visualization, statistical analysis, and machine learning. Such systems provide organizations with a broader understanding of their claim portfolios. The proposed Data-Driven Insurance Claims Analysis system follows this approach by focusing on structured claim-data processing, historical analysis, visualization, comparative insights, and a foundation for future predictive capabilities.

III. System Analysis

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

The Data-Driven Insurance Claims Analysis system is designed to process historical insurance claim records and generate meaningful analytical insights. The system accepts claim data containing information such as policy type, claim amount, claim date, customer segment, location, claim status, and settlement details. The uploaded or connected dataset is first validated to ensure that required fields and supported formats are available. A preprocessing module removes duplicate records and handles missing, inconsistent, or incorrectly formatted values. The cleaned data is then stored in a structured format for analytical processing. The analytics engine calculates metrics such as total claims, approved claims, rejected claims, average claim amount, and total settlement value. The system performs trend analysis to identify changes in claim volume and amounts across different time periods. Comparative analysis can be performed across policy categories, regions, claim types, and customer segments. Visualization components display the results using charts, graphs, tables, and KPI cards. Users can apply filters to examine specific periods, policy categories, or geographical segments. Reports can be generated from the analyzed information for operational and management purposes. The system provides analytical evidence that can support further investigation and decision-making by authorized insurance personnel.

Existing System

Existing insurance claim analysis is often performed using spreadsheets, database queries, and manually prepared reports. Claim information may be stored across different operational systems, making it difficult to obtain a unified analytical view. Employees may need to collect information from multiple sources before preparing reports. Manual data preparation can involve considerable effort, particularly when claim datasets contain thousands of records. Spreadsheet-based analysis may require repeated calculations and manual creation of charts. Identifying claim trends across multiple periods or categories can therefore become time-consuming. Existing reporting systems may provide standard reports but may offer limited interactive analysis. Users may also have difficulty comparing claim performance across regions, policy types, and customer segments. Missing or inconsistent data can affect the accuracy of manually generated reports. Unusual patterns may not be immediately visible when information is presented only in tables. Some systems focus mainly on transaction processing rather than analytical visualization. As a result, organizations can benefit from a centralized system that automates data preparation, analysis, visualization, and reporting.

Disadvantages of Existing System

- Manual data analysis requires considerable time and effort.
- Claim information may be distributed across different systems.
- Spreadsheet-based analysis can be difficult for large datasets.
- Repeated calculations and report preparation are required.
- Limited interactive visualization capabilities.
- Difficult to identify long-term claim trends manually.
- Comparing different policy categories can be time-consuming.
- Missing and inconsistent data can affect analysis.
- Unusual claim patterns may not be easily visible.
- Standard reports may provide limited customization.
- Limited centralized access to analytical information.
- Advanced predictive analysis may not be available.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Proposed System

The proposed Data-Driven Insurance Claims Analysis system provides a centralized platform for processing, analyzing, and visualizing insurance claim information. Users can upload historical claim datasets or connect the system to an authorized data source. The system validates the incoming data and checks important fields such as claim ID, policy type, claim amount, date, status, and settlement information. A data preprocessing module removes duplicate records and handles missing and inconsistent values before analysis. The cleaned dataset is stored in a structured database or analytical data store. The analytics engine calculates important KPIs such as total claims, approved and rejected claims, average claim amount, and settlement values. Trend analysis identifies changes in claim volumes and financial values over time. Users can compare claims across policy categories, regions, claim types, and customer segments. Interactive dashboards present analytical findings through charts, graphs, tables, and KPI cards. Filters allow users to focus on specific dates, regions, policies, or claim statuses. The system can generate downloadable reports for management and operational use. Future machine-learning modules can provide additional indicators such as anomaly detection and claim-risk analysis while keeping final decisions under appropriate human review.

Advantages of Proposed System

- Centralizes insurance claim analysis.
- Reduces manual data-processing effort.
- Handles large claim datasets efficiently.
- Provides automated KPI calculations.
- Identifies claim trends over time.
- Supports policy and region comparisons.
- Provides interactive dashboards and visualizations.
- Makes claim patterns easier to understand.
- Supports filtering and customized analysis.
- Generates analytical reports efficiently.
- Provides a foundation for predictive analytics.
- Helps support data-driven operational decisions.

IV. Methodology

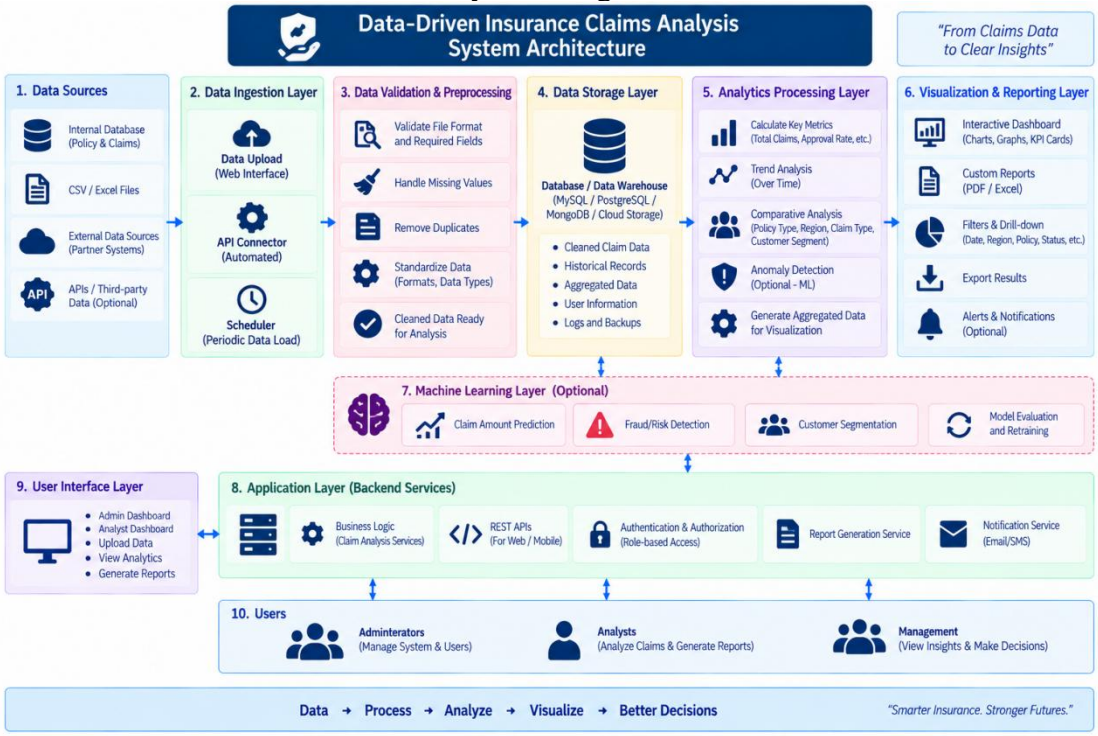
The methodology begins with collecting insurance claim records from an authorized dataset or connected data source. The system validates the uploaded data and checks the structure, required fields, data types, and file format. Data preprocessing is then performed to remove duplicate records and handle missing, inconsistent, or invalid values. Important fields such as claim amount, claim date, policy type, location, status, and settlement amount are standardized for analysis. The cleaned data is stored in a structured database or analytical data store. The analytics engine processes the data and calculates key metrics such as claim count, approval rate, rejection rate, average claim amount, and settlement value. Trend analysis is performed to understand changes in claim volume and financial values across different time periods. Comparative analysis is used to examine differences between policy categories, claim types, geographical regions, and customer segments. Visualization techniques convert analytical results into KPI cards, charts, graphs, tables, and interactive dashboards. Users can apply filters and drill down into specific sections of the claim dataset. The system generates reports containing selected analytical results for authorized users. The complete methodology combines data collection, validation, preprocessing,

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

storage, descriptive analytics, trend analysis, visualization, filtering, and reporting to transform raw insurance data into useful information.

System Architecture

The system architecture consists of multiple interconnected layers that process insurance claim data and deliver analytical insights to authorized users.



The User Interface Layer provides facilities for uploading datasets, selecting analysis options, viewing dashboards, and generating reports. The Data Input Layer receives claim records from files, databases, or authorized external sources. The Data Validation Layer checks file formats, required fields, data types, and basic data quality. The Data Preprocessing Layer removes duplicates, handles missing values, corrects inconsistent formats, and prepares the dataset for analysis. The Data Storage Layer securely stores cleaned claim records and historical analytical data. The Analytics Processing Layer calculates claim KPIs, settlement statistics, approval and rejection rates, and other descriptive measures. The Trend and Comparative Analysis Layer examines claim patterns across time periods, policy categories, regions, and claim types. The Visualization Layer presents results through dashboards, charts, graphs, tables, and KPI cards. The Reporting Layer generates downloadable reports containing selected analytical findings. A Security Layer manages authentication, authorization, access control, and protection of sensitive insurance information. An optional Machine Learning Layer can later provide anomaly indicators, predictive analysis, or risk-related insights for further human review. Finally, the Output Layer delivers understandable claim analytics and reports to authorized users for operational analysis and decision support.

V. Result and Output

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

VI. Conclusion

The Data-Driven Insurance Claims Analysis system provides an efficient solution for analyzing large volumes of insurance claim data and converting them into meaningful insights. The system combines data preprocessing, analytics, visualization, and reporting to reduce the effort involved in manually analyzing claim records. It provides a centralized platform for understanding important aspects of insurance claim operations.

The system analyzes information such as claim amount, policy type, claim status, settlement amount, claim date, customer segment, and geographical region. It can calculate important metrics including total claims, approved claims, rejected claims, average claim amount, and settlement values. This helps users understand the overall performance and distribution of insurance claims.

The dashboard provides visual representations of claim trends, policy-wise comparisons, regional patterns, claim-status distribution, and settlement information. Interactive filters and charts allow authorized users to examine specific periods, policy categories, regions, or claim statuses. These features make large amounts of claim information easier to understand and support evidence-based operational analysis.

The proposed system can be useful for insurance analysts, administrators, and management teams by reducing manual reporting and improving access to analytical information. It can also highlight unusual trends or patterns that may require further investigation. Analytical indicators should be treated as decision-support information rather than automatic determinations about individual claims or customers.

In the future, the system can be enhanced with Machine Learning-based claim prediction, anomaly detection, fraud-risk indicators, automated report generation, real-time data integration, customer segmentation, and advanced forecasting. These

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

enhancements can make the platform more intelligent and scalable while providing deeper insights into insurance claim operations.

References

[1] Kumar, R. D., Prudhviraj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring COVID-19 through intensive investigation with supervised machine learning algorithm. In Handbook of Artificial Intelligence and Wearables (pp. 145-158). CRC Press.

[2] Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November). Machine Learning Techniques in Cloud Based Intrusion Detection. In The International Conference on Artificial Intelligence and Smart Environment (pp. 557-564). Cham: Springer Nature Switzerland.

[3] Sv satyakrishna, shirisha rangu ,bhargavi nalacheruve.(2024) Prospective investigation on colorectal cancer with SMOTE on machine learning Algorithm

[4] Dr.G.Vishnu Murthy, BhargaviNalacheruve 1Professor, Department of computer Science & engineering, Anurag University, TS, India. 2Student, Department of computer Science & engineering, Anurag University, TS, India.

[5] V. N. S. Manaswini, K. K, C. Nigam, S. S. Ali, R. Niranjana, and Suman, "Real-Time Object Detection in Drone Surveillance Using YOLOv5," in Proc. 2025 3rd Int. Conf. IoT, Communication and Automation Technology (ICICAT), Gorakhpur, India, 2025, pp. 1–6, doi: 10.1109/ICICAT68430.2025.11414670.

[6] B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, "Contextual Analysis of Big Data Analytics in Intelligent Transportation Frameworks," in Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart Environment, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer, 2025, doi: 10.1007/978-3-031-88304-0_79.

[7] R. D. Kumar, V. N. S.Manaswini, "Applications of blockchain in smart cities: detecting fake documents from land records using blockchain technology," in Blockchain for Smart Cities, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.

[8] Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, "An advanced movie recommender using collaborative filtering and sentiment analysis," International Research Journal of Modernization in Engineering Technology and Science, vol. 7, no. 7, July 2025, doi: 10.56726/IRJMETS81618.

[9] Ravi Kumar Banoth, Ramana Murthy B V, "Automatic crop recommendation system using LightGBM and decision tree machine learning models," Journal of Machine and Computing, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.

[10] Ravi Kumar Banoth, Dr. B.V. Ramana Murthy, "Smart agriculture through IoT and machine learning for analyzing carbon footprints," in Proc. Int. Conf. Computer Science and Communication Engineering (ICCSCE), Apr. 2025.

[11] Ravi Kumar Banoth, B. V. Ramana Murthy, "Soil image classification using transfer learning approach: MobileNetV2 with CNN," SN Computer Science, vol. 5, art. no. 199, 2024, doi: 10.1007/s42979-023-02500-x.