Incident Pattern Analysis Report:

Industrial Safety Risk Insights

Billy Pierre

---

1. Executive Summary

This report analyzes over 400 industrial workplace accidents across anonymized sites in multiple countries and industries. While most reported incidents were minor (Level I), a significant percentage had high potential severity (Levels IV–VI). Risk analysis and machine learning techniques were used to identify patterns, trends, and actionable insights that can help reduce risk, improve safety, and support strategic decision-making.

---

2. Project Objective

To explore and analyze patterns in industrial accidents and identify:

- High-risk activities and sectors
- Severity trends over time
- Key risk types and their distribution
- Opportunities for predictive prevention

---

3. Dataset Overview

- Source: Kaggle (IHMStefanini Industrial Safety Dataset)
- Records: 400+ accidents from 12 facilities
- Fields: Date, Location, Industry Sector, Risk Type, Severity Levels, Gender, Worker Role, Free-text Description

---

4. Tools Used

- Python (Pandas, Matplotlib, Seaborn)
- Scikit-learn (Random Forest)
- WordCloud for NLP
- Google Colab for analysis and visualization

---

5. Key Findings

Incident Types

- **Top risks:** Pressed, Manual Tools, Chemical Substances
- Over 20% of entries were initially labeled “Other/Unknown” — later cleaned and grouped.

Sector and Location

- **Mining and Metals** reported the highest number of incidents.
- Most incidents occurred in **Country_01** and **Local_03**.

Timing Trends

- Incident volume **peaked in early months** (Feb–June).
- 2016 had nearly twice as many incidents as 2017.

Worker Profile

- **Third-party personnel** faced more incidents than direct employees.
- **Male workers** were involved in 95%+ of cases.

---

6. Advanced Analysis

- A **Random Forest model** was trained to predict whether an incident had high potential severity.
- Input features: accident level, industry sector, gender
- Output: binary classification of “High Severity” (Levels IV/V/VI)
- Results showed that **accident level and industry sector** were strong predictors.

---

7. Business Recommendations

Risk Mitigation

- Prioritize safety audits in Mining and Metals
- Focus training on Pressed, Manual Tools, and Chemical risk exposure

Predictive Strategy

- Use model insights to flag high-risk worker profiles
- Implement early-warning systems using this model's logic

Operational Planning

- Increase safety monitoring in **Q1–Q2**
- Improve incident classification to reduce reliance on "Other/Unknown"

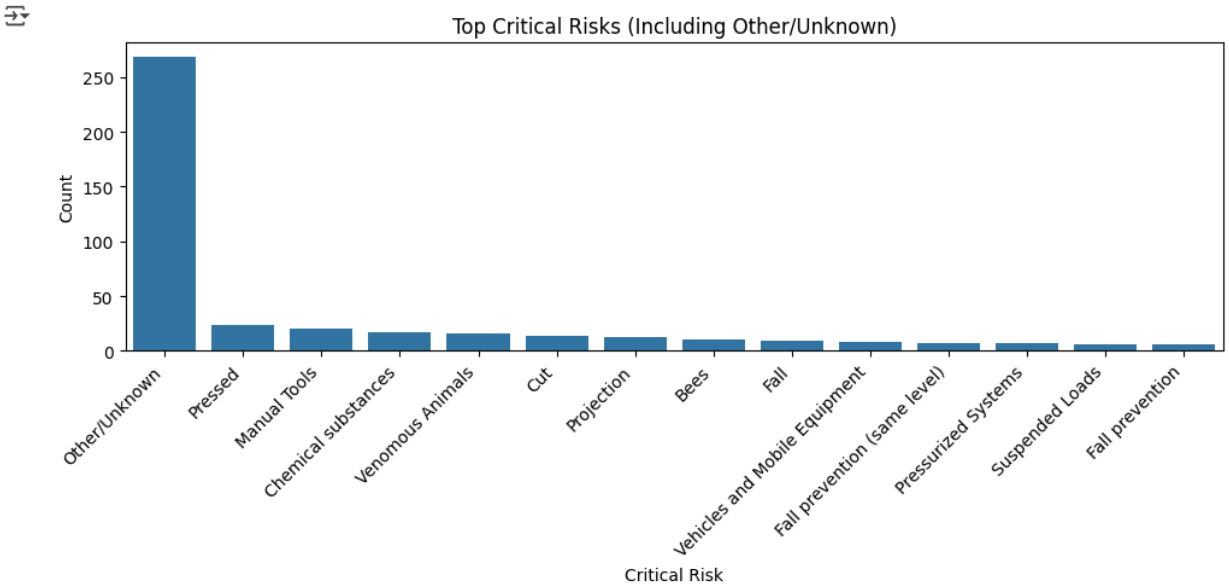
---

8. Limitations & Future Steps

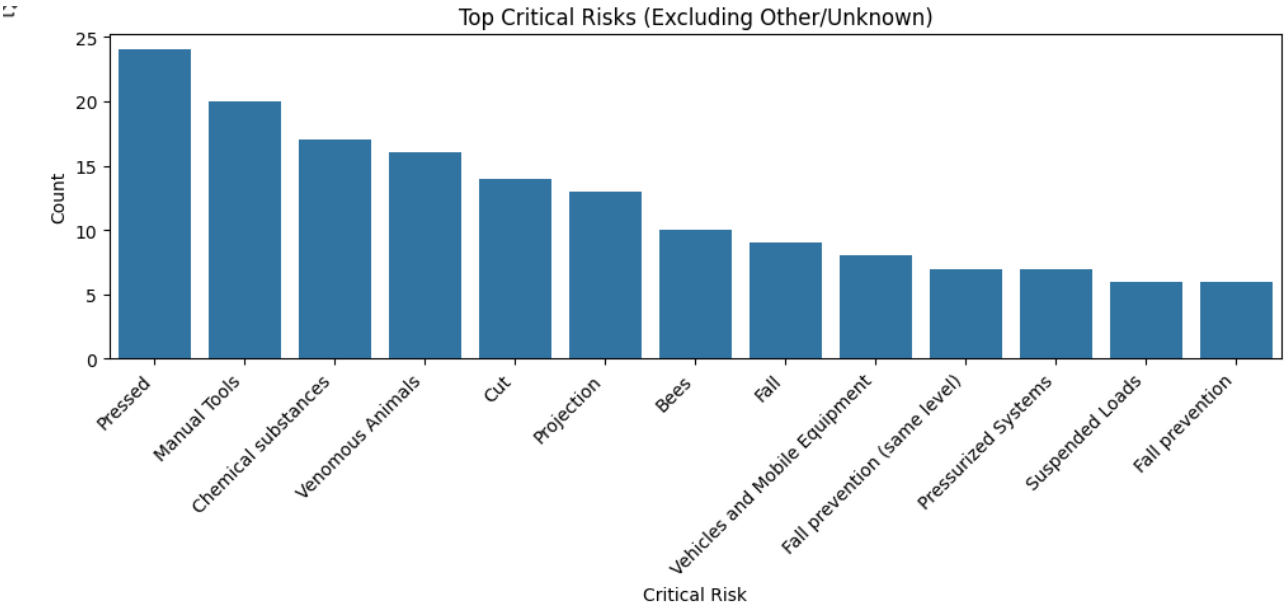
- Limited number of Level VI incidents (only 1) constrained model training
- Descriptions were not fully mined for causation
- Future work could involve:
    - Real-time dashboards
    - Broader NLP to classify descriptions
    - Facility-specific risk scoring

---

9. Visual Highlights

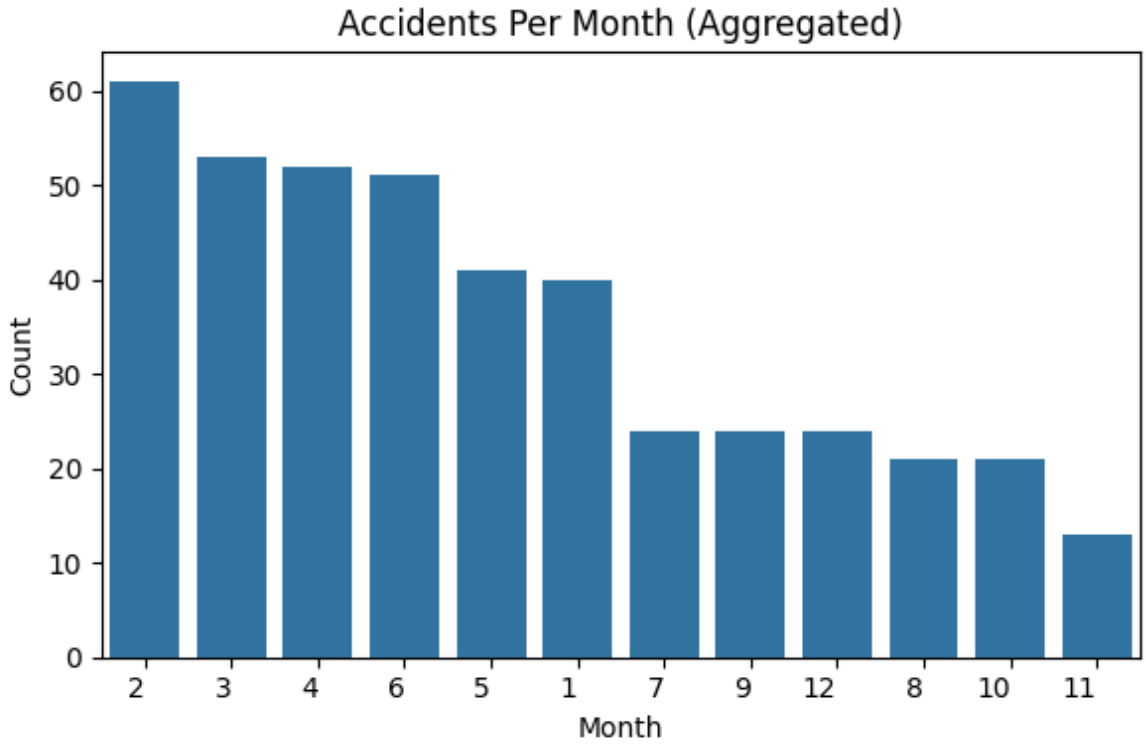


The majority of incidents were classified as "Other/Unknown," indicating poor or inconsistent categorization. This weakens root cause analysis and limits actionability.

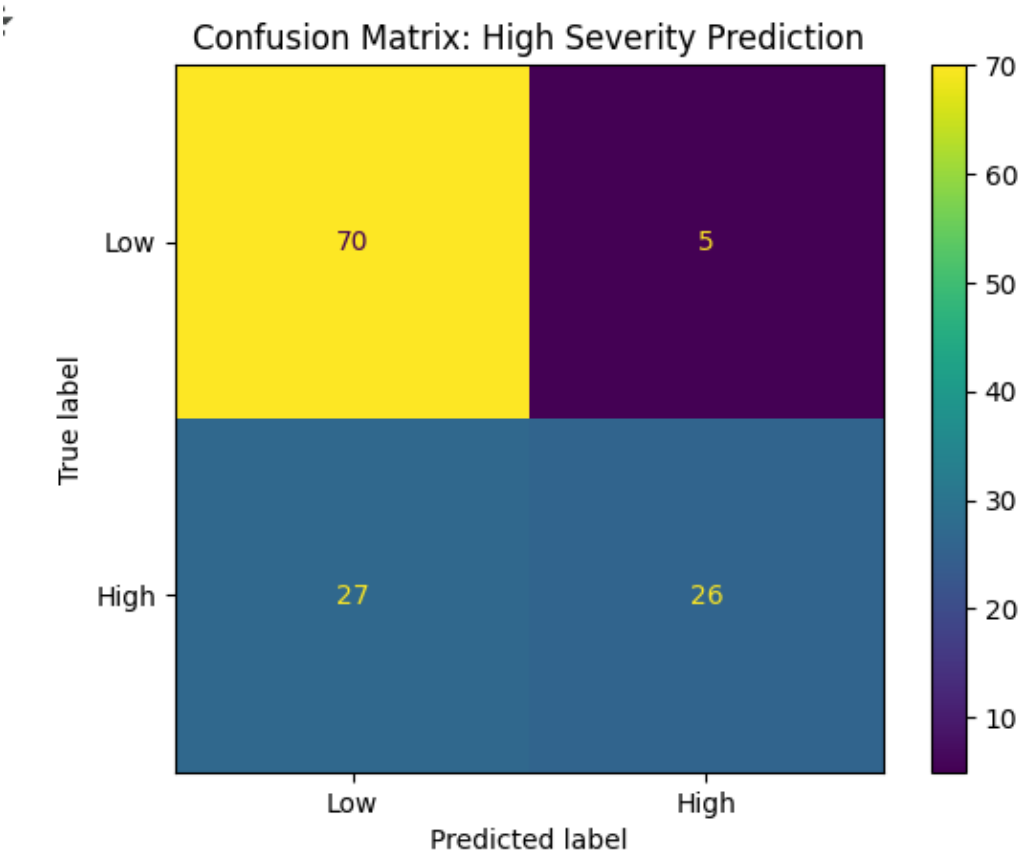


When grouped cleanly, “Pressed,” “Manual Tools,” and “Chemical Substances” emerge as the top risk categories — suggesting where preventive efforts should be prioritized.

Most sectors reported Level I incidents; however, Metals and Mining also had a higher proportion of Level IV–V accidents, indicating latent risk severity in those sectors.



Incident frequency peaks in Q1 and Q2, with February being the most accident-prone month. This suggests a need for heightened safety measures early in the year.



The model performs reasonably well, correctly identifying high-severity incidents 26 times. However, 27 high-severity cases were missed, suggesting room for improvement or more features.

Key terms like “employee,” “causing,” “operator,” and “equipment” suggest that human-machine interactions are frequent contributors to accidents. These terms can guide deeper text mining or training content.