Data Science Solution Summary Report

👤 Name: Vishal Kumar Sharma

📧 Email: vishalsharma270503@gmial.com

📅 Date of Submission: [30-07-2025]

📘 Internship: AICTE Virtual Internship – Data Science with RapidMiner

# 1. Project Title

COVID-19 Global Trend Analysis Using RapidMiner

# 2. Objective

The project aims to perform exploratory data analysis and predictive modeling on COVID-19 data. Using RapidMiner, we analyze confirmed cases, recoveries, and deaths across countries to find patterns, evaluate lockdown impacts, and predict future trends.

# 3. Dataset Used

• Source: Kaggle / WHO  
• Format: CSV  
• Attributes: Country, Date, Confirmed, Deaths, Recovered

# 4. Tools and Techniques

• Platform: RapidMiner Studio  
• Data Cleaning: Handling missing values, removing nulls  
• Visualization: Line plots, bar charts, scatter plots  
• Models: Logistic Regression, Decision Tree  
• Evaluation: Split Validation, Cross-Validation  
• Metrics: Accuracy, Precision, Recall

# 5. Workflow Summary

1. Import Dataset using "Read CSV"  
2. Clean and preprocess data  
3. Visualize trends using plot operators  
4. Create features like Active Cases  
5. Build models and validate  
6. Export insights for report generation

# 6. Key Insights

• Strong correlation between lockdown and infection decline  
• High recovery rates in early responders  
• Logistic Regression showed better accuracy than Decision Tree in this dataset

# 7. Conclusion

The AICTE internship project enhanced practical skills in RapidMiner and demonstrated the ability to solve real-world data science problems. The approach from data preprocessing to model evaluation helped extract meaningful insights from the pandemic data.

# 8. Repository Link (if applicable)

🔗 https://github.com/Vishal-Kumar-Sharma-2003/Data-Science-Solution-Summary-Report.git