### **QUANTUM WOLF**

### **DATA INTELLIGENCE RESEARCH LAB**

Date:08-02-2025-12-02-2025

Name: ABDUL ZUHAIL M

**Automated Misinformation Detection Pipeline Using KNIME**

**1️.Problem Statement**

The rapid spread of fake news and misleading information online has become a significant challenge in the digital age. Misinformation can influence public opinion, create social unrest, and even impact political and economic stability. Traditional methods of fact-checking are time-consuming and cannot keep up with the speed at which misinformation spreads. There is a critical need for an automated, real-time solution to detect and classify misinformation effectively.

**2. Solution Overview**

The proposed solution is an Automated Misinformation Detection Pipeline using KNIME, a data analytics platform. The pipeline:

* Collects data from online sources (social media, news websites, and forums).
* Processes and cleans data using Natural Language Processing (NLP).
* Classifies content as real or fake using AI/ML models (RoBERTa, BERT, GPT).
* Visualizes misinformation trends to monitor its spread.
* Automates processing and reporting using KNIME’s workflow automation features.

**3.Working**

**Step 1: Data Collection**

* **Source**: NewsAPI (**https://newsapi.org**)
* **Query**: Extracts real-time news articles containing the keyword **“misinformation”**.
* **Output**: A structured **Excel file** containing relevant news articles.
  + Columns: Title, Description, Content, Source, Published Date, URL.

**Step 2: Data Preprocessing**

* Cleans and processes text for analysis.
* **Preprocessing Steps:**
  + **Remove Special Characters & Punctuation** (e.g., ! @ # $)
  + **Remove Stopwords** (e.g., “the,” “is,” “and”)
  + **Normalize Text** (Convert to lowercase)
  + **Remove Extra Spaces**
* **Output:** Cleaned text stored in a new **Excel file**.

**Step 3: Fake News Classification**

* **Model Used**: Pre-trained **RoBERTa model** in KNIME.
* **Process:**
  + **Tokenization**: Converts text into numerical format.
  + **Model Inference**: RoBERTa predicts probabilities.
  + **Softmax Activation**: Converts logits to probabilities.
  + **Final Prediction:**
    - **Class 1 (Real News)** – Higher probability → Reliable news.
    - **Class 0 (Fake News)** – Higher probability → Misinformation.
* **Output:** Classified news saved in **Excel format** for trend monitoring.

**Step 4: Data Visualization**

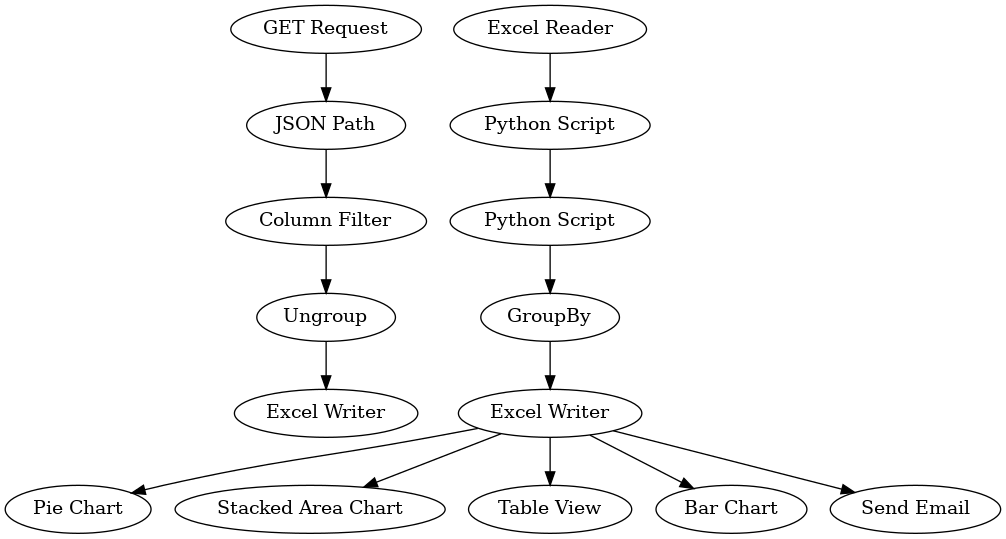
After classification, data visualization helps analyze trends:

* **Fake News Percentage** – Pie Chart (Real vs. Fake News distribution).
* **Top Sources** – Bar Chart (Major sources of misinformation).
* **Geographical Trends** – Stacked Area Chart (Tracking spread of misinformation over time).

**Step 5: Automated Report Generation**

* **KNIME “Send Email” Node** automates reporting.
* Processed results are emailed to stakeholders for review.

**4.Flow Diagram**



**5.Conclusion**

The **Automated Misinformation Detection Pipeline** efficiently classifies and monitors misinformation using AI and KNIME workflows. The pipeline enables **real-time analysis, visualization, and reporting**. Future enhancements may include:

* **Integration with Apache Airflow** for full automation.
* **Expanding data sources** for better accuracy and robustness.

This streamlined process ensures a scalable and effective approach to tackling misinformation in today’s digital landscape.