One Pager Report

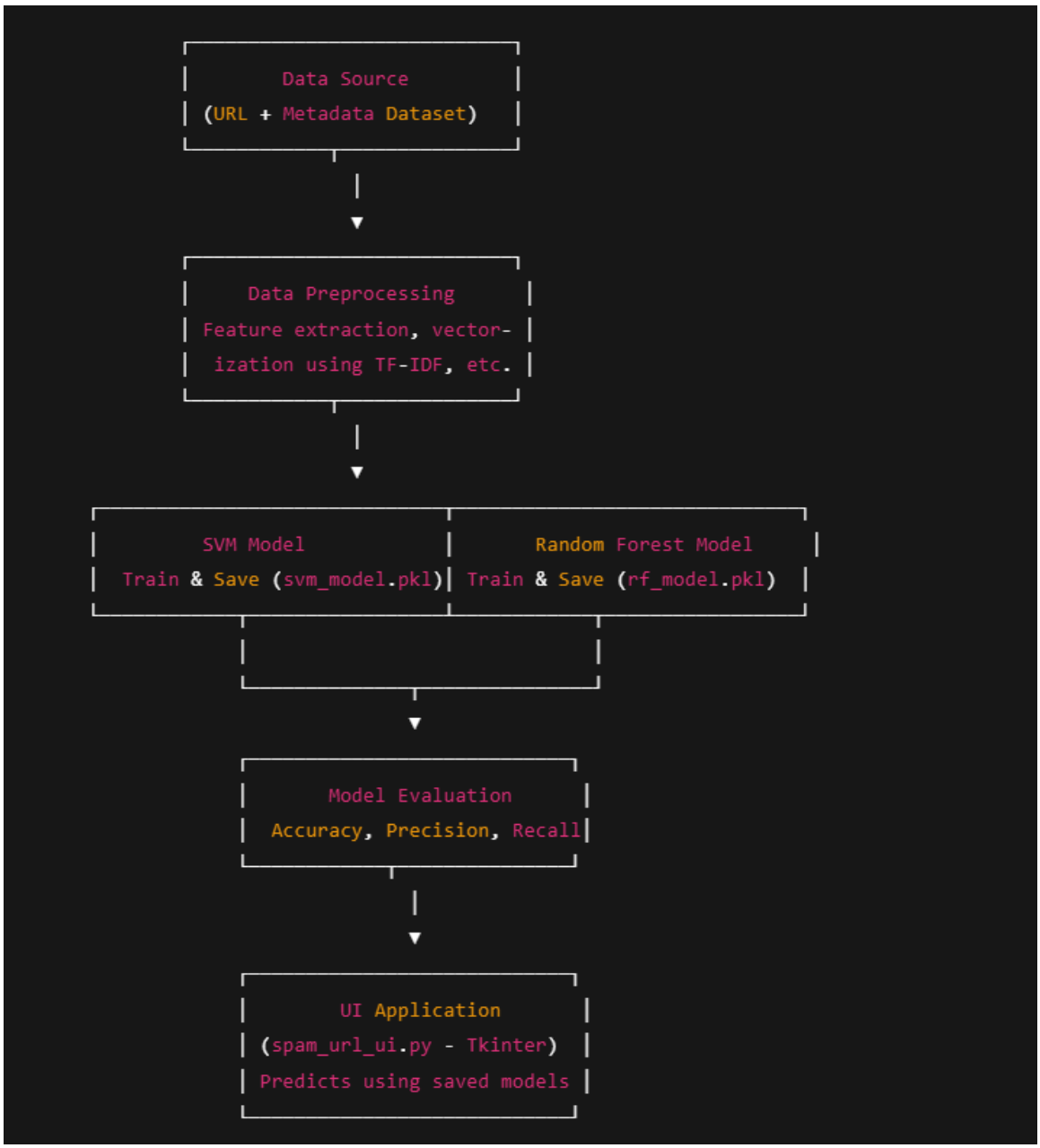
Project Title:

Spam URL Detection Using Machine Learning

Project Statement:

Spam websites waste crawl resources and degrade the quality of search engine results. Traditional detection methods depend on crawling and analyzing page content, which is time-consuming and resource-intensive. This project develops a machine learning-based solution that classifies URLs as spam or safe using only the URL string and metadata (e.g., domain info, registration details, and traffic stats), without accessing web content. By leveraging Support Vector Machine (SVM) and Random Forest (RF) models, the system enables scalable, efficient spam detection.

High-Level Architecture:



Results:

Model	Accuracy	Precision	Recall	F1-Score
SVM	99.28%	0.99	1.00	0.99
Random Forest	98.67%	0.99	0.98	0.98

Both models achieved high accuracy (>98%), ensuring reliable spam detection. The SVM model slightly outperformed Random Forest in precision and F1-score. A simple Tkinter-based UI allows users to input a URL and view predictions from both models in real time.