**Crypto Insight 1.0**

**1. Data Extraction**

* **Input:** Multiple PDF files
* **Method:** Extract text using pdfplumber
* **Output:** Raw text from all PDFs

**2. Text Preprocessing**

* **Steps:**
  + Convert to lowercase
  + Remove special characters
  + Normalize Unicode text
  + Tokenize text
  + Remove stopwords
  + Lemmatize words
* **Tools:** NLTK, SpaCy, Unidecode

**3. TF-IDF Vectorization**

* **Method:** TfidfVectorizer from sklearn
* **Process:**
  + Convert text into numerical vectors
  + Capture word importance across documents
* **Output:** TF-IDF embeddings for sentences

**4. Sentiment Analysis**

* **Method:** TextBlob sentiment polarity
* **Output:** Positive, Negative, or Neutral sentiment
* **Used to adjust BM25 ranking**

**5. BM25 Answer Retrieval**

* **Initialize BM25 model:** Tokenized sentences
* **Search Queries:** Tokenize user input
* **Sentiment Multiplier:** Adjust ranking
* **Output:** Top-ranked answers

**6. User Query Processing**

* **Steps:**
  + Accept user input
  + Analyse sentiment
  + Retrieve & rank answers using **BM25 & TF-IDF**
* **Output:** Most relevant crypto-related answer

**7. Interactive Query Mode**

* **Continuous User Interaction**
* **Exit with "STOP" command**
* **Display:** Answers ranked by BM25 & TF-IDF scores