To establish an accuracy monitoring pipeline for a model, we must navigate through several crucial phases:

Data Collection and Enhancement:

Collecting audio data and accurately annotating it with text is just the beginning. It's essential to augment these audio files with additional metadata such as accent, gender, duration, and age, mirroring classifications seen in Kaggle datasets. This enhanced annotation process facilitates deeper analysis, especially in identifying model drift.

In terms of text annotation, normalizing the text to focus on semantic meaning rather than variations in pronunciation or spelling is key. This means aligning different spellings like "categorise" vs. "categorize," or variations such as "a" vs. "an," and "we're" vs. "we are." Using a script to normalize these discrepancies before adding them to the database streamlines the data handling process significantly.

Data Organization and Performance Indicators:

Before importing the data into an Elasticsearch database, the setup should facilitate ease of evaluation. The Word Error Rate (WER) is a common metric for assessing speech recognition models [1], making it a suitable choice for our purposes. Calculating WER requires knowledge of the total word count in the labelled text and the count of accurately transcribed words. It's beneficial to include these counts in each document's metadata. Since WER is calculated as the ratio of incorrectly transcribed words to the total word count, this metric can be dynamically computed using Elasticsearch's runtime fields, necessitating only the indexing of essential metadata like accent, gender, duration, and age for efficient aggregation.

Continuous Monitoring and Visualization:

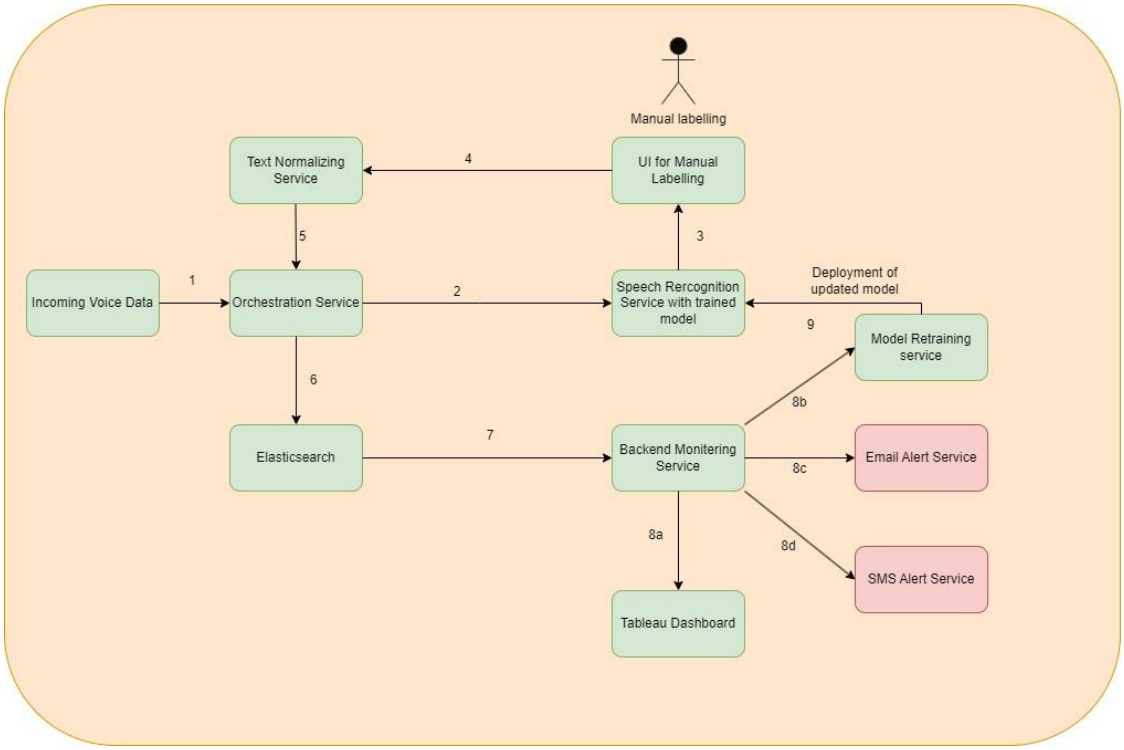
Elasticsearch significantly enhances real-time monitoring by allowing for instant search [2] and aggregation, facilitating detailed tracking of model performance across various dimensions, such as accent, gender, duration, and age. By setting up dedicated dashboards for specific metadata categories, we can efficiently monitor and identify potential sources of model drift, such as the Word Error Rate (WER) among males aged 15 and under.

While Kibana is a reliable and integrated tool for analytics within the Elasticsearch ecosystem, particularly for time series and log analysis, I seek to further enhance the monitoring capabilities. I would like to propose integrating a custom backend server with Tableau for creating dashboards. This combination promises a more dynamic interaction with the Elasticsearch database via API calls, allowing us to leverage Tableau's superior capabilities in producing robust, interactive, and visually engaging insights from complex data sets.

This innovative approach not only enriches our data visualization and analysis capabilities but also extends our alerting mechanisms, from standard email notifications to SMS alerts, thereby broadening our communication channels for reporting critical performance issues or drifts in real time. Additionally, it paves the way for automating essential processes, such as model retraining based on specific performance metrics, ensuring our model remains accurate and relevant. This strategy aims to provide a comprehensive and efficient solution for continuous monitoring and visualization, significantly enhancing our ability to maintain and improve model performance.

478 words

I've designed a diagram to simplify the understanding of our data monitoring pipeline. It details the various services involved in the process. The numbers 1 through 9 illustrate the sequence in which data flows, while labels 8a to 8d represent points in the data flow where processes occur concurrently.



References:

[1] Errattahi, R., El Hannani, A., & Ouahmane, H. (2018). Automatic speech recognitiondata..ors detection and correction: A Review. Procedia Computer Science, 128, 32–37. doi:10.1016/j.procs.2018.03.005

[2] Near real-time search: Elasticsearch Guide [8.12]. (n.d.). Retrieved from https://www.elastic.co/guide/en/elasticsearch/reference/current/near-real-time.html