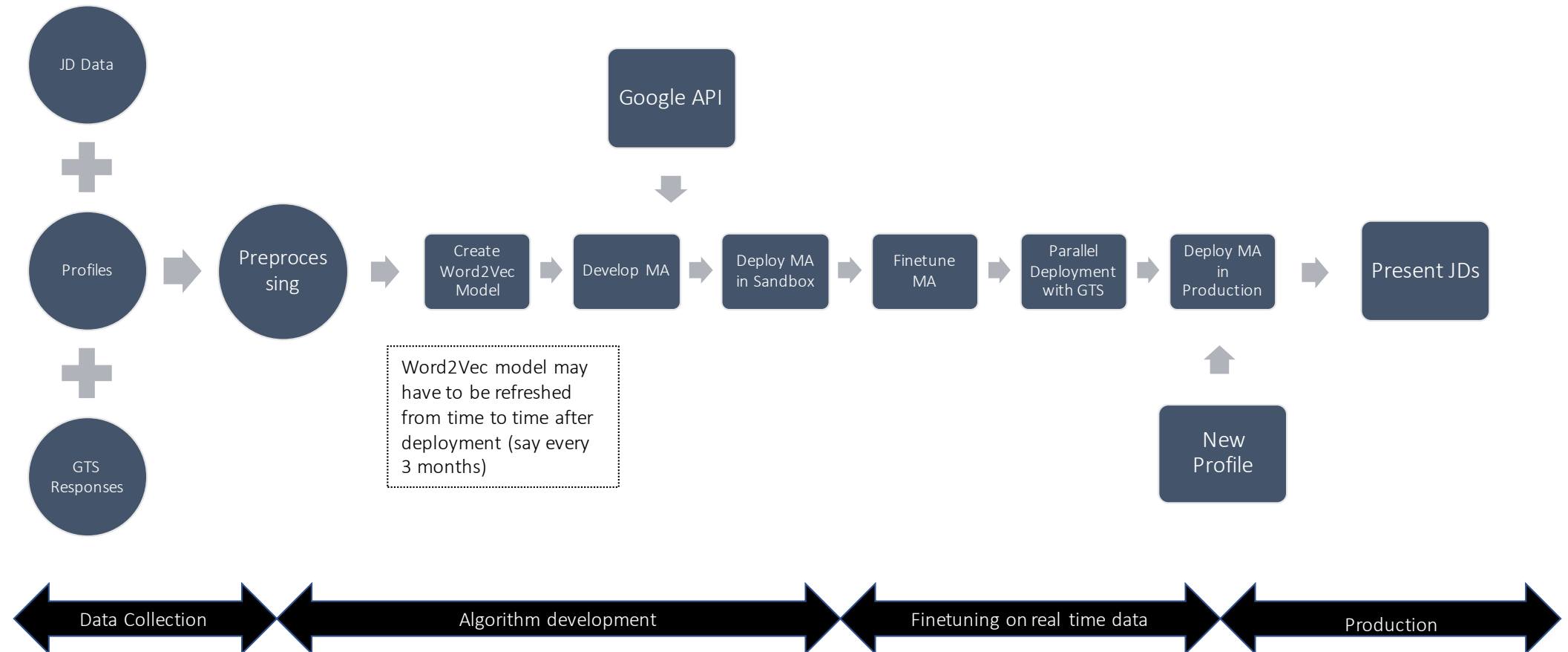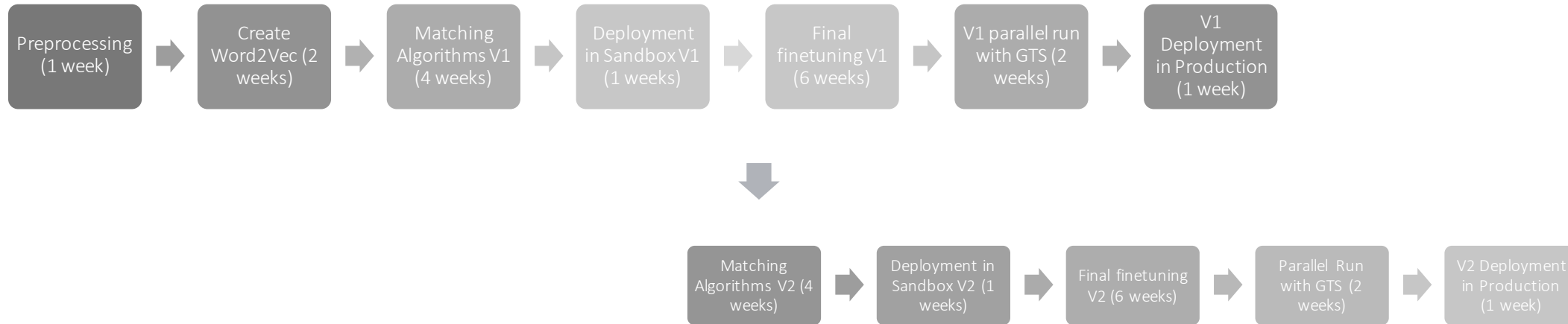# Job Matching Algorithm

9/29/2020

Introduction

- The client has an AI-based matching feature on its platform that finds match between candidate profiles and client job requirements. This feature leverages Google Talent Solution (GTS) algorithm component from Google cloud using its API. However, the GTS is not meeting all Aliro's platform requirements. The project's goal is to build a similar algorithm that fills the current gaps and replace the GTS in the future. Aliro will own the IP and source code and could license it in the future.
- There are about 2 million profiles in client's DB and about 11k active job descriptions which may go up to 22k in a year.

Matching Algorithm: Development and Deployment

JD Data

Profiles

GTS Responses

Preprocessing

Create Word2Vec Model

Google API

Develop MA

Deploy MA in Sandbox

Finetune MA

Parallel Deployment with GTS

Deploy MA in Production

Present JDs

New Profile

Word2Vec model may have to be refreshed from time to time after deployment (say every 3 months)

Data Collection | Algorithm development | Finetuning on real time data | Production

Timeline

Technology Stack

- Python

- Gensim

Data Needed from Client

- Job description data (All or Most)
- Profile data (All or Most)
- GTS response data for comparison

Metrics for Quality of Algorithm

- Precision @ n:
  - Precision is the percentage of selected items that are relevant to the user. For Ranking we can limit it to top n items. Relevance to the user will be as per the results by GTS.

  Precision @ n $= \dfrac{Number of relevant items}{Total number of items (n)}$

- Mean Average Precision (MAP):
  - The Mean Average Precision (MAP) metric tries to approximate this weighting sliding scale. It uses a combination of the precision at successive sub-lists, combined with the change in recall in these sub-lists.

Example of MAP

- Let us assume that we get five items from the model(n = 5). The order along with relevant(R) or irrelevant(IR) tag is as follows:
  - Item1(R),   Item2(IR),   Item3(IR),   Item4(R),   Item5(IR)
  - (R/IR labelling to be done based on the response from GTS)
- Now first we calculate the average precision for all the 5 items
  - Step1 Precision   = 1 (1 item is relevant out of 1).
  - Step2 Precision   = 0.5 (1 item is relevant out of 2).
  - Step3 Precision   = 0.33 (1 item is relevant out of 3).
  - Step4 Precision   = 0.5 (2 item is relevant out of 4).
  - Step5 Precision   = 0.4 (2 item is relevant out of 5).
- Mean Average Precision   = (1+0.5+0.33+0.5+0.4) / 5
  - = 0.546