

Individual Contract

Data Scientist

Background

How can we obtain information about the spread of infectious disease outbreaks in real time? How can we use data from the general population, including those who do not go to doctors or clinics, to enhance existing sentinel systems? How can we combine this data and other data sources into an Early Warning System that allows us to identify early outbreaks and forecast disease spread?

At Pebble Analytics, we strongly believe that adequate responses to emerging infectious disease outbreaks, such as COVID-19, are strongly dependent on the availability of high-quality data about the geographical circulation of the disease and its symptomatic expression in humans and animals.

However, existing health monitoring systems are limited by several factors, including chronically insufficient financial and human resources and limited institutional infrastructure. In addition, going to a clinic or hospital is often an arduous, expensive, and time consuming process. People, particularly those that struggle to afford care or lack insurance, often wait until they are really sick to bring themselves in. During this time they can continue to infect others, which is exacerbated by these individuals often living in higher-risk areas. This reporting delay can be fatal to prevention efforts that aim to stop the spread of a disease before it can cause significant damage. Individuals may even choose not to go to the hospital and recover on their own, which means that their symptoms are never reported and the disease is able to continue to circulate, undetected.

So what if we could leverage the pervasiveness of ICT technologies and the availability of digital, social, demographic and behavioral data to change how traditional disease monitoring systems work?

At Pebble Analytics, we're creating the world's first crowdsourced Early Warning System for outbreaks. We leverage data directly from the general population and combine it with big data sources to offer an alert system that help decision-makers act before it's too late.

Objectives

We are looking for an awesome Data Scientist to apply data mining techniques to collect and scrape the internet for signals about the circulation of outbreaks using structured and unstructured data. The Data Scientist will be responsible for the acquisition, filtering, and analysis of data from at least two sources in order to (i) detect and highlight patterns and trends related to outbreaks; (ii) develop a filter/classification system to determine risks in population health; and (iii) work with our Data Engineers to help link and automate data cycle processes to feed an existing high-performance computing platform.

Key Responsibilities

The Data Scientist will:

- Identify relevant data, using data science methods (e.g., supervised and unsupervised machine learning, NLP, or network science) to scrape and extract information on early signals of potential infectious disease outbreaks.
- Use said methods to query a selected number of global databases (e.g., web-based, open and/or official data), such as Google trends, Healthmaps, global news outlets, reports of animal health, population density data, global infectious disease alert, climate report, or Insect Vectors and Animal Diseases Reservoirs.
- Create a list of keywords associated with health conditions/outbreak risks, employing filters on information from different sources to narrow the results to the issue at hand and using clustering tools that allow it to quickly, and in real time, identify areas or regions with the potential to become hot spots, cold spots, and/or spatial outliers

(priority countries/languages will be provided).

- Design and deploy a secure data lake to store the information gathered and the requirements of the infrastructure in terms of privacy. This should also include data from Digital Participatory Surveillance (i.e., crowdsourced data).
- Define appropriate initial filters and cleaning criteria to treat incoming data (big and crowdsource data), including data validation procedures, missing value treatment, etc., needed for ingesting, storing, and using the data.
- Pre-process the data coming from different sources, regions, and formats in order to standardize it and
- Define the proper communication channels for presenting the current and forecasted status and the calculated impact to targeted clients.

Priority Countries

Kosovo, Côte d'Ivoire are priority - potentially adding South Africa, Kenya, Mozambique.

Duration

Initial contract (part-time) is for 3 months with possibility of extension and growth

Location

Remote or at our sunny offices in Barcelona

Qualifications

We don't like putting people into boxes and we look for talent and the potential to learn and grow quickly. Having said that, here are a few qualifications to guide your application (but don't be shy about them!)

Pebble Analytics

- Degree in Computer Science, Computer Engineering or related technical discipline.
- Experience in Data Analysis/Software Engineering.
- Experience with Big Data analytics (NLP, Network Analysis, ...) and applying these to solving real-world problems, an asset.
- Experience with professional software engineering practices and best practices for the full software development life cycle, specifically:
 - Coding capabilities in Python or R and Data Science related libraries (
 - Experience with analytics & visualization tools
 - Experience with cloud computing platforms such as GCP, AWS, Azure or others
- Ability to communicate complex concepts in a clear and effective manner.
- Experience performing analysis with large datasets
- Desire to work in a highly-collaborative environment is an asset

Contact

Natalia Adler (natalia@pebbleanalytics.com)