

Nationality Prediction of any Individuals by their First Names Using Machine Learning

Sakil Sarker (2017-2-60-079)
Sayeda Fatema Tuj Zohura (2017-1-60-011)
Md. Abuhorayra (2017-2-60-077)

September 10, 2020

1 Abstract

Is it possible to predict nationality? To answer questions about ‘Ethnicity or Nationality’, we often look at individuals names first. The majority works until now has been conducted considering the last name. But in most of the cases, the first name holds much information and really matters a lot. To make a relationship between first name and personality, we collected some names from internet sources and labeled them with their desired nationality. In this project, we developed a machine learning model to deploy a system that can predict the Nationality of individuals from their first names. Here we developed different types of models and then tested them with machine learning techniques to identify the best model among them.

2 Introduction

Nationality is, a people sharing a common origin, culture, and/or language, and possibly constituting a nation-state ([Your Dictionary](#)). Nationality is also a legal identification of a person in international law, establishing the person as a subject, a national, of a sovereign state. It affords the state jurisdiction over the person and affords the person the protection of the state against other states ([Wikipedia](#)) . Name analysis is often the only practical way to gather nationality annotations because there are also privacy concerns. Yet, nationality remains one of the fundamental attributes that can not be easily masked like age and gender even in disguise. In this project, we collected data contain first names and corresponding nationalities through which we will try to predict an individual’s nationality which was not in the data set. To do this, we used supervised learning which actually infers a function from labeled training data consisting of a set of training examples. We tried to make nationality prediction work much accurate so it can accurately label the nationality of an individual.

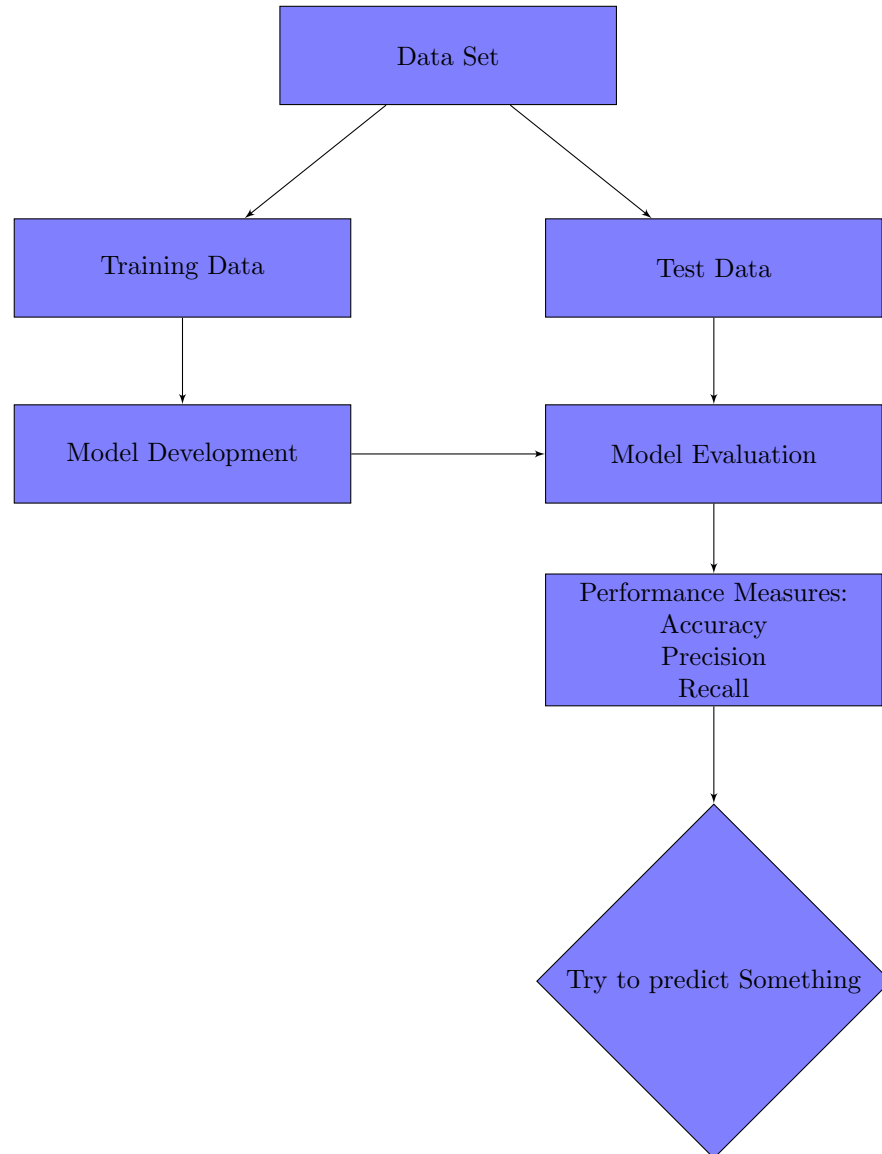
3 Objectives

1. To collect an average numbers of data from different nationality.
2. To create a model and fit those data to the model.
3. Run machine learning algorithm.
4. Check whether our model can predict nationality or not.

4 Motivations

1. Nationality prediction can be a key component that can be deployed in various systems at security checkpoints where they have the list of names [1].
2. Nationality identification may unlocks important demographic information, with many applications in biomedical and sociological research [1].
3. From social media posts and comments, by applying nationality prediction may demonstrate stark differences in the nationality of the followers of public figure such as Trump and Obama, and in the sports and entertainments favored by different groups.
4. By applying this, e-commerce businessman's can make targeted add.

5 Model Description



References

- [1] J. Ye, S. Han, Y. Hu, B. Coskun, M. Liu, H. Qin, and S. Skiena, “Nationality classification using name embeddings,” in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pp. 1897–1906, 2017.