

# Ethnic Fractionalization Using Names

## 1 Introduction

## 2 Data

## 3 Methods

### Features

- Split names into n-grams using characters
- Test using all lowercase letters and marking which letters begin and end a name (relevant for naive bayes and SVM, which are order agnostic). Idea here is to basically give algorithm information as to whether n-gram of characters starts a name, ends a name, or is in the middle of a name.

### Algorithms

- Naive Bayes
- SVM
- RNN