ALGORITHMICS FOR DATA MINING

ADM - ASSIGNMENT 2

April 25, 2018

Fabio Di Francesco
David Sarda

Contents

1 INTRODUCTION

We are big python fans so we decided to try one of the libraries for machine learning that are available, *sklearn*. To do so we chose a dataset that looked interesting, it's about homicide reports on the United States between 1980 and 2014. The main columns are as follows:

City/State/Year/Month: Time and place of the incident.
CrimeType: The type of crime commited, in this database it can be either a murder, or a manslaughter by negligence.
Crime solved: It states whether the crime was solved or not (by 2016).
Victim_sex/Victim_age/Victim_race: Characteristics of the victim.
Perpetrator_sex/Perpetrator_age/Perpetrator_race: Characteristics of the main perpetrator.
Relationship: The relationship between the victim and the perpretator.
Weapon: The weapon used for the crime.
Additional_Victim_Count: The number of victims killed by the perpetrator in addition to this one.
Additional_Perpetrator_Count: The number of perpetrators that collaborated with the main perpetrators to execute the crime.
Record_Agency: The agency that was responsible for investigating the case.

2 DATA MINING ALGORITHMS

Naive Bayes

Decision Tree

K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) with PCA Finally, we decided to run a PCA on our

3 FINAL THOUGHTS