Assignment 4

Name: Uday Kumar Kamalapuram

GMU Id: G01340201

Bias in Machine Learning:

AI systems learn to make conclusions based on training data, which may contain distorted human decisions or historical or social injustices. Another type of bias is faulty data sampling, which occurs when some groups are disproportionately represented or underrepresented in the training data. Machine learning is used to make many decisions with business implications, such as loan approvals in banking, and personal repercussions, such as diagnostic conclusions in hospital emergency rooms. Machine Learning bias will have an impact on people's daily life. However, the criminal justice system is currently seeing some of the most concerning instances of prejudice in machine learning.

Machine learning bias is frequently caused by issues brought by those who build and/or train machine learning systems. These people might create algorithms that reflect unintentional cognitive biases or real-world prejudices. Individuals may also add biases by training and/or validating machine learning algorithms using incomplete, incorrect, or biased data sets. Stereotyping, bandwagon effect, priming, selective perception, and confirmation bias are examples of cognitive bias that can mistakenly alter algorithms.

In current project I have observed following biases by running given projects and followed the given steps in document to set up the running environment for the project.

Part 1:

flowers insect pleasant unpleasant (twitter WEAT file)

Effect Size : 1.25

When I run the WEAT file of twitter with these insects and flower with pleasant and unpleasant attribute. I got the Effect Size of **1.53** which is positive score, and it represented the bias in the names of flowers. The name of flowers is most towards the pleasant than the names of insects.

When I run the same on Wikipedia generated WEAT file I got Effect size of 1.08

Part 2:

For my part 2 experiment in this Assignment, I have chosen the city names from two different continents one is North America and Africa. I wanted to check whether there is bias in these city names, as from the general public view there is bias of cities in Africa treated as unpleasant places, whereas the cities of North Americas biased as the cities of dream, happy and healthy.

I have chosen thirty city names from the North America mostly from USA, I have chosen top city names from USA and some other top city names from Canada. The following is the word list.

City_names_America: los angeles, new york, san diego, san francisco, washington dc, dallas, chicago, houston, phoenix, philadelphia, san jose, san antonio, austin,fort worth,jacksonville, charlotte, columbus, boston, las vegas, miami, arlington, Jersey City, hollywood, kansas city, toronto, vancouver, ottawa, quebec, montreal, calgary.

I have chosen thirty city names from the Africa mostly from these countries Central African Republic, Libya, Somalia, South Sudan, Mali.The following is the word list.

City_names_Africa: Bangui, Birao, Nola, Bria, Bouar, Tripoli, Benghazi, Misrata, Derna, Tobruk, Sirte, Ghadames, Mogadishu, Hargeysa, Berbera, Bosaso, Gaalkacyo, Juba, Wau, Yei, Malakal, Aweil, Kuajok, Rumbek, Yambio, Bamako, Sikasso, Kalabancoro, Koutiala, Kayes.

NorthAmericaCityNames AfricaCityNames pleasant unpleasant (twitter WEAT file)

Effect Size : 1.08

When I run the WEAT file of twitter with these NorthAmericaCityNames and AfricaCityNames with pleasant and unpleasant attribute. I got the Effect Size of **1.08** which is positive score and it represented the bias in the names of two continents. The City names of North America is most towards the pleasant than the city names of Africa continent.

The top 5 similar words generated for NorthAmericaCityNames are cheer, sunrise, crash, vacation, family.

The top 5 similar words generated for AfricaCityNames are murder, filth, tragedy, prison, pollute.

NorthAmericaCityNames AfricaCityNames pleasant unpleasant (wiki WEAT file)

Effect Size : 1.55

When I run the WEAT file of wikipedia with these NorthAmericaCityNames and AfricaCityNames with pleasant and unpleasant attribute. I got the Effect Size of **1.55** which is positive score, and it represented the bias in the names of two continents. The city names of North America are most towards the pleasant than the city names of Africa continent.

The top 5 similar words generated for NorthAmericaCityNames are crash, friend, health, honor, sunrise.

The top 5 similar words generated for AfricaCityNames are pollute, caress, agony, filth, sunrise.

Conclusion:

From running both twitter and Wikipedia WEAT files I have observed that there is a bias introduces on city names based on the Continents where it belongs. It may affect the decision making of Machine Learning on the various applications of people coming from these cities or may directly affect the people who are leaving there.