

Project 2

Bilguun Chinzorig, Monika Avila,
EPFL

Abstract—

I. INTRODUCTION

The aim of this project is to

II. THE MODEL

III. METHODS

IV. THE DATA

The training data set is formed by 2 millions of tweets, half of them correspond to the positive emotions and the other half to negative ones. This data has been already tokenized, thus we start with the cleaning and the analysis of the data.

First, we remove the hashtags, URL directions and @...

Second, we counting the words and we check Zipf's law.

After this, we extract features: ... a. n-grams b.

V. RESULTS

VI. DISCUSSION

VII. SUMMARY