# Introduction

Big companies, such as Facebook and Google, know more of us than we can imagine. They do not only store our browsing history, but also, the time we spend online, the links we click and our mouse trajectories. This data is then processed to find our interests, among other things, so that they can provide us personalized advertisements.

I am attracted by the strength of marketing on social media. Social Media Marketing is a relatively new approach to Social Marketing. Social Marketing was introduced as a concept in the early 70s. Kotler and Zaltman (1971) used this concept to refer to the promotion of social objectives, such as brotherhood and safe driving. This interpretation of social marketing became less important and when you ask a marketer how they interpret social marketing nowadays, they will explain that social marketing is related to applying marketing through social networks, online or offline. It should be noted that, although my study will focus on purchase intent, that Social marketing is not only related to trying to sell as many products through social networks. In general, Social marketing is used in applying all 4 Ps that are known in the marketing mix: product, price, place and promotion (Thackeray et al., 2008). There exist different purposes of social marketing: viral marketing, increase buyer’s loyalty, increase product awareness. In this study, I will focus on the latter and I am going to find out if it is possible to efficiently identify purchase intents without involving all the other data sources (e.g. mouse trajectories, browse history) which companies like for example Google and Facebook deploy in their approaches. It would be too simplistic to assume that big companies only filter on relevant keywords and show advertisements related to the keywords you used in your posts. In their approaches more than keyword filtering is used to ensure that undesirable scenarios are prevented. Such a scenario could for example be that someone receives a promotion for cigarettes after he posted a tweet (a micropost with a maximum of 140 characters on Twitter) about cigarettes in a negative context: “@username: No more cigarettes #diagnosis #lungcancer”. Even so, there still exist cases that seem innocent in which advertisements are not filtered, but these cases can nevertheless be very painful for the person itself. A simple example of this would be a woman seeing advertisements of baby clothes 9 months after she started following Facebook pages with information about pregnancy, although she has now stopped searching this because of an unwanted abortion. Such could have been avoided with an approach understanding more of the personal context1.

To work on this problem, I am going to combine existing techniques (e.g. sentiment analysis2) with my own functions to extract features from tweets. Then, I will create a model that tries to predict whether or not someone can be a potential customer. This prediction will only be based on streams of short messages (i.e. the tweet history of a person). I will apply the model to tweets of people that use product-related terms in one of their tweets. The model can be trained using labels that are manually assigned by an employee of the marketing division of the company that is used. This employee determines whether the author of tweets is a potential customer (PC) or not (non-PC) in his opinion. Besides, the annotation, she marks the number of tweets that had to be read before the decision could be made. After a model is trained, it is applied to a subset of the data. How these subsets are formed is explained later. I can evaluate the model by looking at the precision3 and recall4. The evaluation should give an answer to the question: To what extend can AI approach the predictions (potential customer vs. no potential customer) that are assigned by a human expert?

Some of these posts may reveal their potential purchase intent towards something (e.g. “What phone would you recommend me to buy?”). Identifying an intention of purchase from these posts is useful, and at the same time challenging. It is useful for enhancing community observing systems using social data (Miluzzo et al. 2008), for improving the knowledge gathering process of context-aware systems (Banerjee et al. 2009; Sakaki, Okazaki, and Matsuo 2010) and for targeted marketing solutions.

In concept, we define a Purchase Intention (PI) as a text expression signifying an intention to purchase or consume a product or a service. Any post not having such an expression belongs to non-PI category. Specifically, our goal is to solve the two-class classification task - is the post a PI or a non-PI? Therefore, our focus is on extracting important features having significant discriminating power to enable an effective classification of PIs from user-posts. There are several characteristics of short-texts which pose a challenge to this task:

* Informal language: Language used in social media is often informal and lacks grammatical structure. Usage of acronyms and miss-spellings is also common. These factors not only make the semantic analysis of the text difficult, but also make the data high dimensional in nature.
* Limited contextual information: The short size of posts in social media (e.g. 140 characters limit on Twitter; our dataset had an average of 73 characters per post) makes it difficult for the traditional knowledge extraction algorithms to work effectively (Han, Pei, and Yin 2000). It is fairly difficult to gather contextual information from such short texts.
* Language processing-based ambiguities: Multiple possibilities in syntactic parsing of texts, ambiguities in resolution of anaphoric expressions and other linguistic ambiguities make it difficult for natural language processing-based feature extraction systems to work effectively (Zukerman and Litman 2001).

In this paper, we take into consideration these characteristics of short social-posts to design an efficient classification model for detecting PIs. This is done by extracting features at two different levels of text granularity - word and phrase-based features and grammatical dependency-based features. This technique is based on the fact that even though these short posts lack grammatical structure, but at a sub-sentence level, users tend to arrange words in correct order (Banerjee et al. 2012). Apart from these, the patterns observed in PI posts help us to identify some specific features.

Our project is a web application that can predict the likelihood/certainty that a customer will buy a product that he is interested in based on his social media posts such as Twitter tweets and user profile data. This will help the company/business target that particular customer more efficiently and boost its sales.

We will be searching for Twitter tweets of potential customers wanting to buy a product. And based on those tweets we will estimate/predict the likelihood that the customer will buy the product. We will make a model by gathering tweets from users who have already expressed intention to buy the product and see their tweet history and if possible, their web search history as well. Using this model, we will input potential customers who have tweeted about the product but have not bought it. And based on the training data the model will estimate a prediction/likelihood of whether the customer will buy it or not. We will limit the scope of our data to only mobile phones. And our model will predict the consumer intention for the latest upcoming mobile phones by gathering data and updating the model at fixed intervals. We plan to generate a trending list of phones on twitter which will be launching soon or have been recently launched and the users can use that list of products on our application to find a customer for that product/mobile phone.

# Literary Review

Several researches have been performed for analyzing the insights of online consumers buying behavior. However, only a few have addressed the customers buying intention for products. Studies on identification of wishes from texts, specifically Ramanand et al. (Ramanand, Bhavsar, and Pedanekar 2010) consider the task of identifying ‘buy’ wishes from product reviews. These wishes include suggestions for a product or a desire to buy a product. They used linguistic rules to detect these two kinds of wishes. Although rule-based approaches for identifying the wishes are effective, but their coverage is not satisfactory, and they can’t be extended easily. Purchase Intention detection task is close to the task of identifying wishes in product reviews. Here we don’t use the rule-based approach, but we present a machine learning approach with generic features extracted from the tweets.

For our research we will be using twitter tweets to identify purchase intention. Past studies have shown that it is possible to apply Natural Language Processing (NLP) and Named Entity Recognition (NER) to tweets (Li et al., 2012) (Liu et al., 2011). However, applying NER to tweets is very difficult because people often use abbreviations or (deliberate) misspelled words and grammatical errors in tweets. Nonetheless, Finin et al. (2010) tried to annotate named entities in tweets using crowdsourcing. Other studies used these techniques to apply sentiment analysis to tweets. The first studies used product or movie reviews because these reviews are either positive or negative. Wang et al. (2011) and Anta et al. (2013) analyzed the sentiment of tweets filtered on a certain hashtag (keywords or phrases starting with the symbol that denote the main topic of a tweet). These studies merely analyze the sentiment of a tweet about a product after the author has bought it. We will however be extracting features from tweets to find whether the user has shown purchase intention towards the product or not.

More recently, research articles like *Identifying Purchase Intentions by Extracting Information from Tweets* ( February 8, 2017, RADBOUD U NIVERSITY NIJMEGEN) and *Tweetalyst: Using Twitter Data to Analyze Consumer Decision Process* (The Berkeley Institute of Design) investigate if an artificial intelligence approach can predict (from existing user created content on twitter) if someone is a potential customer for a specific company or product and identify users at different stages of the decision process of buying a given product. Further looking at research reports like *The Impact of Social Network Marketing on Consumer Purchase Intention in Pakistan: Consumer Engagement as a Mediator* (Asian Journal of Business and Accounting 10(1), 2017) give us an insight of the impact of social network marketing on consumer purchase intention and how it is affected by the mediating role of consumer engagement. Based on UGT theory (Uses and Gratification Theory).