1.1 Clarify which LDA we are going to discuss. 1-2 Sentence abt the LDA we r not going to talk

The abbreviation term LDA may mean two completely different machine learning algorithm -- Linear Discriminant Analysis and Latent Dirichlet Allocation. Linear Discriminant Analysis is the method used to to find a linear combination of features that characterizes or separates two or more classes of objects or events[[1]](#footnote-0). This research paper will be focusing on the Latent Dirichlet Allocation - a probabilistic-based Topic Modeling algorithm.

2.1 Definition  
 “Topic modeling is a form of text mining in order to identifying patterns in a corpus ”.   
 “A method for finding and tracing clusters of words (called “topics” in shorthand) in large bodies of texts ”.   
 The term “topic” here has been defined by Twitter as a topic as “a recurring pattern of co-occurring words.”  
 Topic models are algorithms for discovering the main themes that pervade a large and otherwise unstructured collection of documents. Topic models can organize the collection according to the discovered themes .

2.2 Usage  
 The primary usage of topic models is analyze text. By applying them into a collection of documents. The latent topics can be discovered beneath the large amount of words, as well as the relationships between different topics, even the evolvement of these topics. The study of Science magazine uses topic model (LDA) to extract 100 topics among 17,000 papers.   
 Also, since the idea of topic can be extended beyond text, topic models can also be improved and modified for other purposes. Models can be adapted to many kinds of data. Among other applications, they have been used to find patterns in genetic data, images, and social networks. 1more example in other topics

2.3 Characteristics of Topic Models  
 Most Topic Models fall into the category of clustering data mining skills. They are also unsupervised learning since they act to reduce data dimension without specifying the values/characteristics of each targets.

However, there are also supervised topic models. By pairing a response to each document in the process of modeling, supervised topic models can be used to infer latent topics predictive of the response.

2.4 Algorithms  
 There are a variety of algorithms that has been developed for topic modeling. The most basic and widely used model is Latent Dirichlet Allocation, which is evolved from Latent Semantic Allocation. Based on LDA, there are also Supervised Latent Dirichlet Allocation .

Most topic models which are currently used evolve from LDA. Part of the models are listed as below:

1. https://en.wikipedia.org/wiki/Linear\_discriminant\_analysis [↑](#footnote-ref-0)