Introduction needs to be a chapter

* Include challenges, contributions, and stories
* Even the related works needs to be well grounded and deeper

Dataset should not be a chapter

Justification for any ungrounded argument

Dataset belongs to testing for hypothesis

Report on the filtering

1. 我们最开始的问题是什么 是模型没法改 没法快速研究多种模型 所以我们用其他方法 然后选择一种特定的东西去研究 为什么选择这个来研究 有很多不同的东西可以做 为什么最后做了那个 每做一种方法发现了什么问题 于是解决这个问题最好的方法是什么 于是答案是什么
2. Community是网络用户 对于每个机器学习的东西要进行论证 这个方法是用来做什么的 为什么值得做 做出来结果怎样 想要怎样

General structure of thesis

1. What was the initial question
2. Why do we want to use probabilistic programming
3. There are multiple things to research, why did we pick the ones we picked
4. What was the first thing that we explored?
   1. Up side
   2. Down side
   3. Because of downside what else did we do
5. Markov Model – Is not the only factor. Not easy to expand to multiple factors.
6. Probabilistic programming
   1. Needs proper definition of the model
   2. Not too expandable using pymc3
7. Regression
   1. Bad performance
   2. Data has several useless features
   3. Logisitic vs. Ridge
   4. Model has limited number of parameters to fit the data. Not possible.
8. Neural nets
   1. Easy to increase hidden nodes to create more parameters to better fit data
   2. Easy to expand over multiple factors
   3. Good performance
   4. Mainly a classifier
   5. Does not distinguish user classes apart
   6. Gains intuition on correlation, but loses intuition on causation
9. Clustering
   1. Finally distinguishes user classes apart
   2. Loses information in separating data into classes

