Interim Report Plan

3035237470 Li Jun

1. **Project objective and key features**

Implement a real-time customer manage system on cloud.

Key feature I:

Real-time analysis into the big data of customer records and real-time

training (online training on streaming data) to improve recommendation and sales prediction.

Key feature II:

Applicable for different fields of datasets and dynamic switch to suitable algorithms according to data or user’s preference. Applicable for graph streaming input from camera.

Key feature III: improve recommend and prediction algorithm with deep learning algorithms. (after midterm)

1. **System Overview**

Data Storage

* 1. **Technical Parts**

File system: HDFS

Database and analysis tool: Elasticsearch

Data process and visualization: Logstash and Kibana

Machine learning tool: Spark with MLlib, GraphX models

Online training tool: Spark with Spark streaming model

**3.2 Key feature I**

**3.3 Key feature II**

**3.4 Key feature III**

**4. Implementation plan: show the weekly schedule.**

Time schedule:

week 1(Nov 28 - Dec 4):

Explore the details about SVD algorithm, [Collaborative filtering](https://link.zhihu.com/?target=http%3A//en.wikipedia.org/wiki/Collaborative_filtering), Content-based Recommendation algorithm and methods of applying them with Elasticsearch and Spark.

Resources: Machine Learning in Action, Machine Learning with Spark

week 2(Dec 5 - Dec 11):

start to implement data analysis on Elasticsearch and recommend system on Spark. Apply different recommend algorithms with Spark MLlib model and compare their results.

week 3(Dec 12 - Dec 18):

keep Implementing data analysis on Elasticsearch and recommend system on Spark. Apply Spark streaming and GraphX models.

week 4(Dec 19 - Dec 25):

Complete system implementation and apply some interfaces for the system. Test and debug for it.

week 5(Dec 26 - Dec 31):

Write and complete midterm report.