

CSCI946 Assignment

Yao Xiao
SID 2019180015

September 29, 2020

1 Phase 1: Discovery

1. Identify data source.
2. Confirm team members and roles.
3. Check project sponsor's approach.

1.1 About Data

The data for the project should be classified into two categories:

1. Five years of ideas submitted by internal innovation competitions.
2. Notes and records representing innovation and research activities from all over the world.

1.2 Hypotheses

The hypotheses for the project should be grouped into two categories:

1. Descriptive analysis of what is happening to further stimulate related creativity and collaboration.
2. Predictive analysis can provide relevant administrative staff with future investment recommendations, such as how much to invest at what time, etc.

2 Phase 2: Data Preparation

1. IT department set up an analytics sandbox
2. Discovered that certain data needed conditioning and normalization and that missing datasets were critical.
3. Many names were misspelled and problems with extra spaces.

3 Phase 3: Model Planning

1. Use social network analysis techniques to look for innovators.
2. Identify the right milestones to achieve the goals.
3. Trace how people move ideas from each milestones toward the goal.

4. Tract ideas that die and others that reach the goal.
5. Compare times and outcomes using a few different methods.

4 Phase 4: Model Building

1. NLP on textual descriptions.
2. Conduct social network analysis using R and Rstudio.
3. Developed social graphs and visualizations.
4. Verify some of the initial hypotheses.

5 Phase 5: Communicate Results

1. Study was successful in identifying hidden innovators.
2. CTO launched longitudinal studies to track innovation results.
3. This project promoted knowledge sharing related to innovation and research.
4. Also enable the organisation to cultivate additional IP.
5. Communicate one of the key findings.

6 Phase 6: Operationalize

1. Deployment was not really discussed.
2. Key findings:
 - Need more data in future.
 - Some data were sensitive.
 - A parallel initiative needs to be created to improve basic BI activities.
 - A mechanism is needed to continually reevaluate the model after deployment.
 - Informed new investment decisions.