Project Proposal ALY

6080

Team 1

Presented by

Jingjie Wang

Sunny Saini

Esha Palav

Shivam Chaudhary

Introduction- After completing the three annotated bibliography I compiled all of them to have a draft plan of how the project is going to be done. The annotated bibliography explained about how the MVP can be created, what approach will we take, what issues we can face and various real-life examples.

The companies task management team have a vision to create an MVP for emotional tracking within the organization to understand if it is a good day or bad day in the company and report it to the top management in the organization. After performing a successful manual research for two months in the organization they are now looking to create an automated working project which will be helpful within the organization and also serves as a product for external launch in the market.

Analysis- We talked about the necessity of understanding the emotional state of the organization. The organization is run by its employees and if the employees are not happy and satisfied, the efficiency of the whole organization suffers. It is a very clear example of cause and effect on various levels. Optum's top management wants to know if it is a good day or bad day in the company for the very same reason. By knowing the emotional state of the employees, the colleagues and management can share empathy with the people who are having bad day. This empathy will give a feeling of belongingness to the employee as they will realize that the company cares about them and understand what phase they are going through. This will ultimately boost the morale of the employees and increase the efficiency.

We also discussed about the various emotional triggers for employees like-

1. Personal Life Issues
2. Job Satisfaction
3. Teammates behaviour and communication
4. Downsizing, delayering and layoff in the organization

Prospective employees would not love to work in a cold, rational machine (Fineman, 2000; Giardini and Frese, 2004; Schreyögg and Sydow, 2001). A good environment is as much important as the efficiency of the organization. The culture is the face of an organization and for new employees, interested candidates, business partners, potential clients; it is important to have that vibrant culture.

Then, we also discussed the approach we will take for executing this project. The product will work on text mining and sentiment analysis for emotional tracking to understand the emotion felt by the user and provide a positive reinforcement. This product will help keep track of the behaviour and emotional state of the employees within the organization.

The three main challenges with text mining for emotional tracking through Natural Language Processing are-

1. Statement Polarity (Positive, Negative, Neutral)
2. Statement Strength
3. And the biggest challenge, complexity of language

The methodology is divided into three parts with sub-sections as follows-

1. Data Extraction
2. Data processing (pre-processing, emotion annotation)
3. Experiments (Feature extraction, Classification)

We will use **Python** to test, train and code our API. **Tableau or PowerBI** can be used for daily visualization reporting to the top management.

The approach will be simple Likert scale emotion detection with values from 1 to 5 for Good day, Bad day and 1 and -1 for Yes and No.

Our model was constructed regards to the results of nine questions from 16 participants, therefore the model was able to predict a new participant's emotion based on his or her own results of the same nine questions. The model has a predicting accuracy of 94.33% on new data. However, we founded that the model can be improved by making it sensitive to time variable. In other words, we can add the day of week as a new predictor and re-construct the model. According to our common sense, people probably will have a good emotion on Friday rather than Monday. This goal can be achieved by accomplishing the following tasks:

- Convert time data from text to date format
- Calculate which day of week based on the given date information
- Store the day of week data in the database and re-construct the model
- Predict with new model and re-evaluate model's performance
- Able to read time information when a new participant enters the survey, and calculate current day of week
- Set up a constraint that only allow participant take the survey during week days

The date information that was stored in the original dataset was in text format, which was consist of numbers and slashes. Even human being can read "6/17/2019" and consider this string as time information, however, the computer system will not be able to read it as time information. In order to let the system to access to the time information, we need to transform the text string into a pre-defined date format in system. Thereafter, we can let the system get the time information and calculate the day of week.

After got all the day-of-week information, I found another tough problem. All the surveys were taken during weekdays from Monday through Friday, and this makes our database biased. What if a new participant takes the survey on Saturday? Our model won't be able to answer that because the model only knows the results from Monday to Friday. This problem

inspired me that I should set up a constraint that only allow people take the survey during weekdays, in this way the model will always be confident to give response, and it will never receive strange data. Therefore, I designed an if clause inside the while loop to determine whether the survey should begin or stop according to current day-of-week. For example, if a participant opened the survey on Friday, he or she will be able to do the survey, however, if a participant opened the survey on Saturday or Sunday, the program will print out a sentence says, "Today is weekend, take a break!" and stops the survey.

The dashboard that I created is shown below
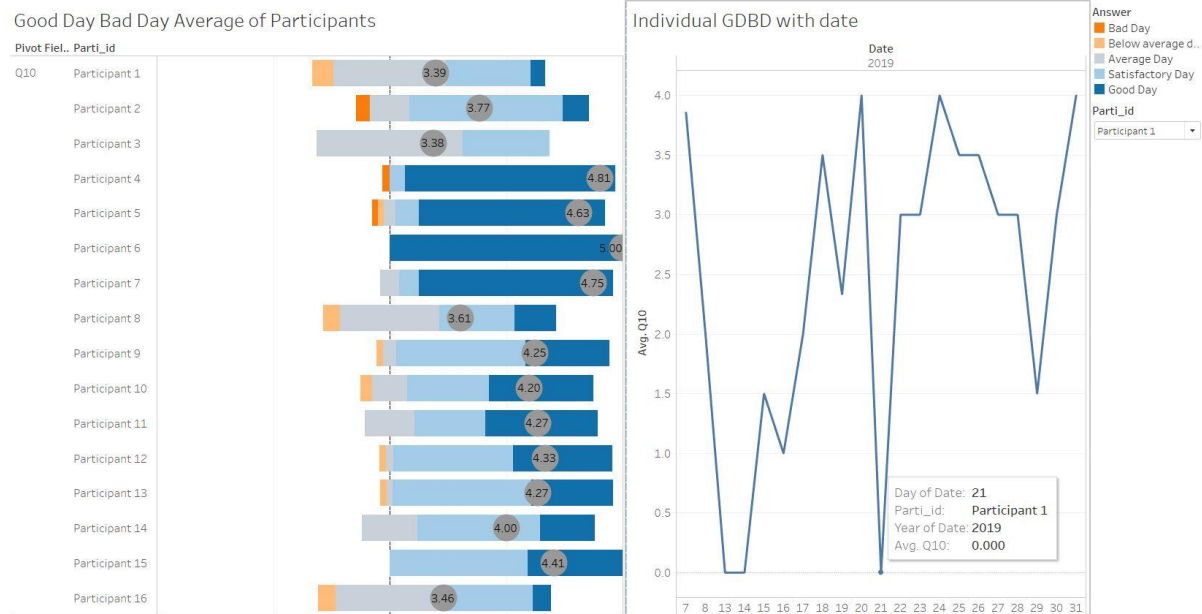
The dashboard contains a Gantt chart which can be used by the top management to view the performance of the whole team and a line chart which shows the performance datewise of individual participant with the filter to change the participant

The calculations I used for creating the dashboard is given below-

Negative Score :-
IF [Pivot Field Values] < 3 THEN 1
ELSEIF [Pivot Field Values] = 3 THEN 0.5
ELSE 0
END

Percent of total sizes:-
SUM([Number of Records])/[Total Scores]

Total Negative Scores:-
TOTAL(SUM([Negative Score]))

Total Scores:-
TOTAL(SUM([Number of Records]))

Gantt Start :-
-[Total Negative Scores]/[Total Scores]

Gantt Percent :-
PREVIOUS_VALUE([Gantt Start]) + ZN(LOOKUP([Percent of total Sizes],-1))

For the Front End part we had decided to build:

1.) A Survey page or the Individual Feedback page
2.) Visualization page or a dashboard in Tableau which would be sent to the Manager so

that he can get further insights.

3.) An Administration Control page that would be accessed by the Manager.

4.) A login page for the Employees to sign in only.

But after downloading the Employee login page which was a Bootstrap plugins we further faced an issue connecting that page to a Tableau dashboard.

Therefore after digging deeper we found out that this could be implemented in Tableau itself:

1.) In the Tableau Desktop, open the workbook you want to publish.

2. Select **Server** > **Publish Workbook**.

If the **Publish Workbook** option does not appear on the **Server** menu, make sure a worksheet or dashboard tab is active (not the Data Source tab).

3.) In the **Publish Workbook** dialog box, select the project, enter a name for the workbook, and add search tags.

Tags help users find related workbooks when they browse the server. Separate tags using a comma or space. To add a tag that contains a space, put the tag in quotation marks.

4.) For **Permissions**, accept the default project settings.

Generally, a site administrator manages permissions on the server. If you think your workbook is an exception, work with your administrator to determine the best course of action, and see [Set Permissions as You Publish a Data Source or Workbook](#).

5.) For **Data Sources**, select **Edit** if you want to change whether the data is embedded in the workbook or published separately, or change how people authenticate with data sources.

☐ If you're publishing an extract and want to set up a refresh schedule, you must select **Embed password** or **Allow refresh access**.

☐ If your workbook connects to a Tableau data source, we recommend embedding the password. If you instead choose to prompt users, they'll need additional permissions on the data source.

For more information, see [Set Credentials for Accessing Your Published Data](#).

6.) Configure [Variable publishing options](#) that are available for this workbook.

7.) Click **Publish.**

8.) (Optional) Set up a refresh schedule for each extract you published.

> The publishing workflow guides you through these steps. For some data types you publish to Tableau Online, the publishing process starts Tableau Bridge on your computer.

After performing these steps now we need to Set Credentials for Accessing the Published Data:

The type of authentication to your data source is independent of how people sign in to your Tableau Online or Tableau Server site. For example, to give people direct access to the data in a workbook, you would embed a database user's credentials into the data source's connection. But anyone viewing the workbook would still need to be able to sign in to the site on Tableau Online or Tableau Server to open your workbook.

The following are the steps that describe how to set authentication on data connections as part of the publishing process.

For many types of connection you can embed a database user's name and password, or use single sign on (SSO). Specific exceptions are described later in this topic.

The following steps describe how to set authentication as part of publishing a data source or workbook. You can do this for each connection in the data source.

1. In the Publish Workbook dialog box, go to the **Data Sources** area, which lists the workbook's connections, and select **Edit**.

2. In the **Manage Data Sources** popup, after you decide whether to publish the data source separately or as part of the workbook, select an authentication type for each connection in the data source.

The available authentication types depend on the connection type, and they can include one or more of the following:

- **Prompt user**: Users must enter their own database credentials to access the published data when the view or workbook loads.

- **Embedded password**: The credentials you used to connect to the data will be saved with the connection and used by everyone who accesses the data source or workbook you publish.

- **Server run as account**: A single Kerberos service account will be used to authenticate the user. On Windows this is the account that Tableau Server runs as. On Linux it can be any Kerberos account.

- **Viewer credentials**: The viewer's credentials are passed through to the database using SSO (usually Kerberos).

- **Impersonate with embedded account** or **Impersonate with server Run As service account**: Impersonation using embedded credentials connects with the

embedded credentials and then switches to the viewer's identity (only for databases that support this). Impersonation using the Run As service account is similar but first, connects with the Kerberos service account before switching to the viewer's identity.

- o **Refresh not enabled** or **Allow refresh access**: These options appear when you publish an extract of cloud data such as from Salesforce, and database credentials are needed to access the underlying data. **Allow refresh access** embeds the credentials in the connection, so that you can set up refreshes of that extract on a regular schedule. Setting **Refresh not enabled** prompts users when they open the workbook.

Important: How you want to keep extracted data fresh is also a factor:

- ☐ If you want to set up an automatic refresh schedule, you must embed the password in the connection.
- ☐ If you're publishing a cloud data connection to Tableau Online, the publishing steps will alert you if you need to add Tableau Online to the data provider's authorized list.
- ☐ You can't publish an extract that's created from a Kerberos-delegated, row-level-secure data source.

Dropbox, OneDrive connections

For Dropbox and OneDrive, when you publish a data source or workbook and select **Embedded password**, Tableau creates a saved credential and embeds it in the data source or workbook.

When you publish a workbook that connects to a Tableau Online or Tableau Server data source, rather than setting the credentials to access the underlying data, you set whether the workbook can access the published data source it connects to. Regardless of the original data type, the choice for server data sources is always **Embedded password** or **Prompt users**.

If you select to prompt users, a user who opens the workbook must have **View** and **Connect** permissions on the data source to see the data. If you select embed password, users can see the information in the workbook even if they don't have View or Connect permissions.

Conclusion- The MVP will be helpful for the top management to keep track of company's culture, emotional state and better understanding of what can be changed in the company to make it better. Knowing the emotional state of company is one of the most essential aspect of business and it can overall improve the efficiency and productivity of the organization.

It is a basic API to collect data in a more structured manner and can be further used for creating an intelligent bot.

References-

1. Küpers W. , Weibler J. (2008) Emotions in Organization: an integral perspective .
    Retrieved from
    https://s3.amazonaws.com/academia.edu.documents/37303599/IJWOE_Emotion_in_
    Organisation.pdf?response-content-
    disposition=inline%3B%20filename%3DEmotions_in_Organization_-
    _An_Integral_P.pdf&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-
    Credential=AKIAIWOWYYGZ2Y53UL3A%2F20200204%2Fus-east-
    1%2Fs3%2Faws4_request&X-Amz-Date=20200204T071612Z&X-Amz-
    Expires=3600&X-Amz-SignedHeaders=host&X-Amz-
    Signature=9e46c0db42a1c898104f73fe189d2b5fc4859cb79742a3e457f645a03c3d01
    48

2. Ranganathan J., Tzacheva A. (2019) Emotional Mining in Social Media Data.
    Retrieved from https://webpages.uncc.edu/jrangan1/papers/KES19.pdf