Paciolan:
Custom Fields

Presented By: Yihan Wang, Sirui Hu, Yadi Yang

Project Summary

- **Background**
  - **Paciolan** is a company aiming to serve the **business clients**, like the university sports and venues, and provide a better platform for them to serve their **patrons and donors (users)**.

- **Goal**
  - To derive business insights in helping define user categories and user potential interests from the **5 client-defined major custom fields** assigned to each user.

- **Objectives**
  - To find out similar characteristics among users who have been assigned common custom keywords/tags.
    - Features will be processed in a **data analysis pipeline**
  - To automatically generate conventional keywords/tags for shared-characteristic users. **(if time allows)**
    - For current users: help improve the tagging process
    - For new users: may not apply because of limited data

1.

Proposed Technical Approach

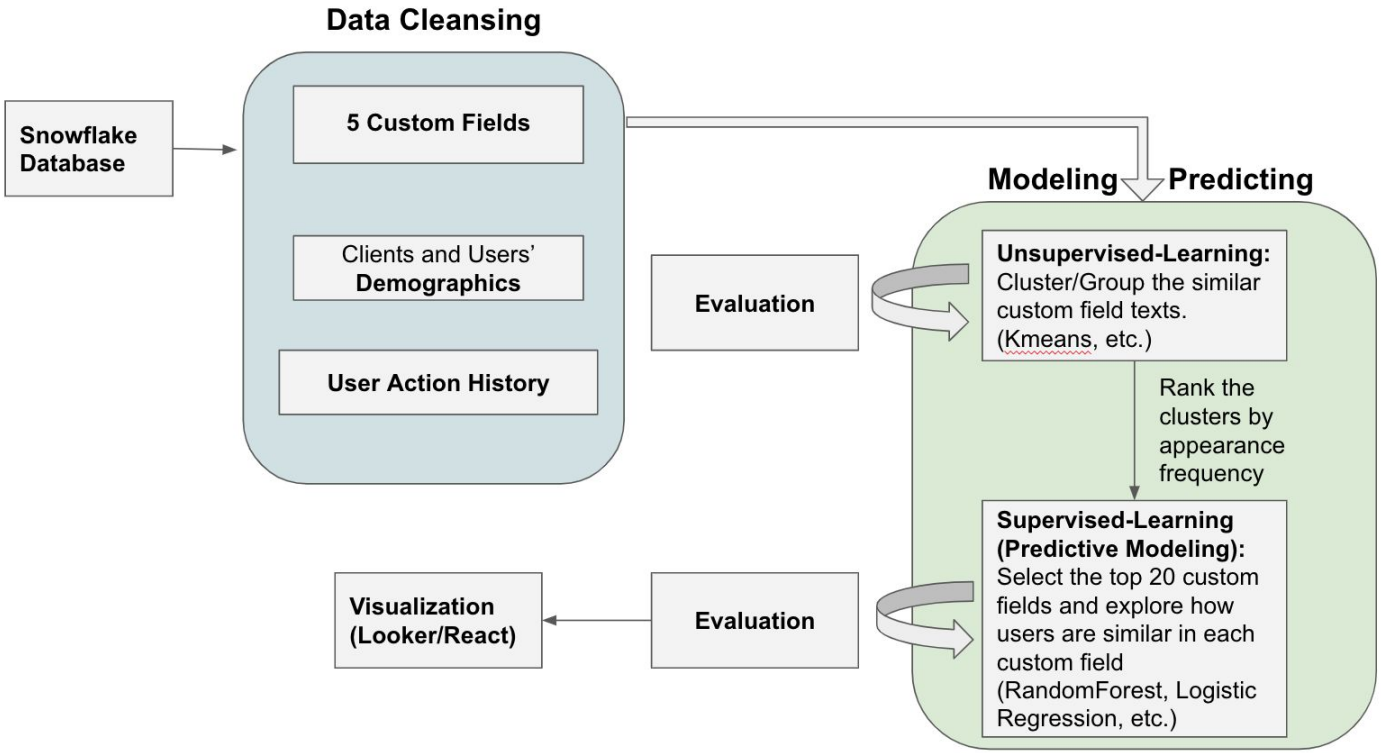
How We Will Break Down the Data

Data including All Clients (Paciolan)

Clients Chosen by Paciolan (Organizations)

Custom Fields Chosen by Organizations (patrons/donors)

Data Analysis Pipeline



2.

Data Sets

Data Sets (Snowflake DB & Json)

1. 5 Custom Fields

- Keywords
- Tags
- User Defined Fields
- Donor Categories
- Motives (Campaign)

2. Clients' & Users' Demographics

- Account Dataset
  - First Three Custom Fields
- Donor Dataset
  - Fourth Custom Field
- Donor Membership Dataset
- Organization Dataset
- Allocation Dataset

3. User Action History

- Transaction Item Dataset

First Three Custom Fields

1. Keywords
Paciolan users' segment accounts into lists for marketing

```
"keywords": [
        "MARTIL",
        "P9999999999",
        "PL4@UCI.EDU"
        ]
```

2. Tags
Paciolan's NumPyent accounts into lists for marketing

```
"tags": [
        "F16REN",
        "F16LINKEDBAL"
         ]
```

3. User Defined Fields
Create new data points on account records to track additional data

```
"userdefinedfields": [
        { "value1": "CPP",
          "value2": "0"  },
        { "value1": "APP",
          "value2": "0"}
         ]
```

Demographic Data - first 3 Datasets

1. Account

User (patron/donor) account information linking to a client

{
“Accountid”: “7102394”,
“Organization_id”: “USC”,
"keywords": ["MARTIL"],
"userdefinedfields": [
{ "value1": "CPP",
"value2": "0" }],
"tags": ["F16REN"]
}

2. Donor

Inheritance of Account. Basic donor information.

{
“Accountid”: “7102394”,
“Organization_id”: “USC”,
"Years_of_donating":24,
“Lifetimedonationamount”: 100000,
“Donorcategories”: {
"2015FBRoadList": "",
"FB18": "1"}
}

3. Donor Membership

Which donor membership a user has with the client

{
“Accountid”: “7102394”,
“Organization_id”: “USC”
“donor_membership” :
{"SPF(MEMBERSHIP_ID)::2012(MEMEBER_DRIVE_YEAR)":
{"pledgedlevel":
{"en_US": "Ralph Young Club"},
"receiptlevel":
{"en_US": "Ralph Young Club"},
"totaldonated": 20000,
"totalpledged": 20000 }
}

Demographic Data - next 2 Datasets

4. Organization Dataset

The dataset stores information about each organization's basic information.

{
"Organization_id": "WISCONSIN"
"Description": "University of Wisconsin - Madison"
"Time_zone": "America/Chicago"
}

5. Allocation Dataset

The dataset stores all the possible places where the users' money may go when an organization receives a payment.

{
"Organization_id": "WASHINGTON"
"Allocationid": "SOCCRM"
"Name": "Mens Soccer Husky Sports Fund"
"Groupname": "Team Enhancement"
}

Action History - Transaction Items Dataset

Detailed Transaction History

- Who sent the money (account id)
- Who received the money (accountdbid)
- How much is the money (paymentamount)
- Where the money will be allocated (allocationid)
- Which motive (custom field) the transaction is related to if any (comments)


```
{
    "accountdbid": “MSSTATE”,
    "accountid": "799439",
    "allocationid": "BDC",
    "comments": "BDC - GENERAL DONATIONS",
    "motiveid": "CR", (one of the custom fields)
    "paymentamount": "25000"
}
```

3.

Experiments and Evaluation

Experiments & Evaluation

- **Unsupervised Learning (Cluster Custom Fields)**
  - Manually Checking whether the clustering make sense
  - Feedback from Paciolan

- **Supervised Learning (Binary classification)**
  - 5-Fold Cross Validation
  - Precision and Recalls
  - Feedback from Paciolan and from their clients

4.

Sources/Softwares

Public Available Sources:

- **Database**
  - Snowflake
- **Compiler/IDE**
  - Jupyter Notebook
- **Present Final Result**
  - Looker (snowflake) or React.
- **Repository**
  - GitLab/GitHub
- **Communication:**
  - Jira (Agile project management)
  - Confluence, Slack, Zoom
- **Packages:**
  - **Snowflake connector: snowflake.connector**
  - Data Wrangling: numpy, pandas, json
  - Data Analysis: scikit-learn, scipy, statsmodel
  - Data Visualization: matplotlib, seaborn, plotly

Personal Available Sources:

- **Programming Language:**
  - Python
  - R (Optional)

- **Compiler/IDE:**
  - Visual Studio Code
  - Jupyter Notebook
  - R Studio (for statistical analysis only)

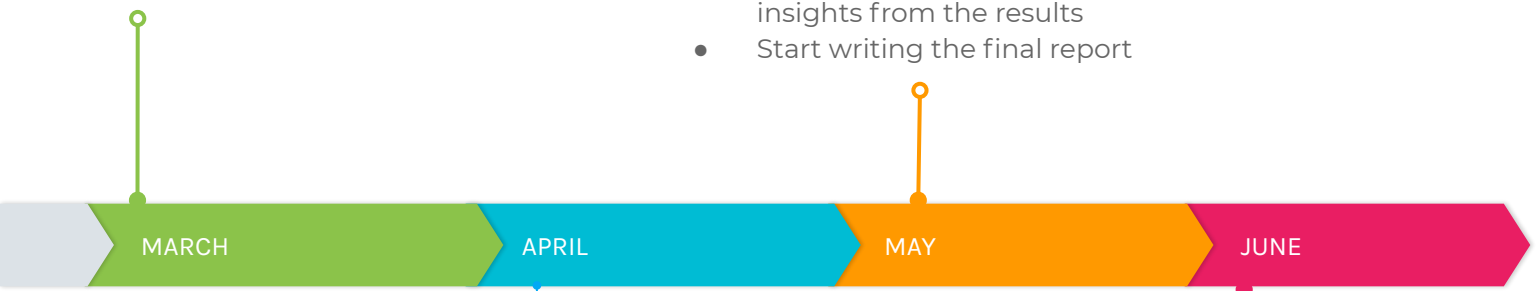
5.

Milestones

Milestones

- ☑️ Set up accounts
- ☑️ Explore the raw data
- ☑️ Draft the proposal presentation.
- Research relevant literature and methods to come up with some initial ideas.

- Continue brainstorming and experimenting on any possible approaches.
- Evaluate the results of experiments.
- Use Looker/React to build data visualization for the deliverable results
- Automate the process of assigning tags and gain insights from the results
- Start writing the final report



- Set up a weekly agenda and recurrent zoom meetings with the Paciolan team.
- Preprocess the raw data and start cleaning and organizing the data for future use.
- Discuss with the sponsor and know their preference

- Optimize results and prepare for the final presentation
- Finalize the final report

THANKS!

Any questions?