HACKEREARTH TEAM NAME - PRACHI570_F599

VIL CODEFEST IDEA SUBMISSION

AI Powered Customer Profiling Solution

VYOM AGARWAL 🔗

IIT BHU VARANASI

https://github.com/vyster/VIL_CODEFEST

VIL CODEFEST
PROBLEM STATEMENT

AI Powered Customer Profiling Solution :-

Exploring how extracted VIL data features can be leveraged for Creating Customer Profiling Solutions through ML/AI techniques.

SOLUTION APPROACH

After noting points mentioned in the PS :-

For creating a solution we can start by an ideation of making **Knowledge Graphs and Taxonomies** between entities (CUSTOMERS). Focus can be on how these and other heterogeneous data sources fit into a flexible architecture that can further build a high level Architecture to create Taxonomies and Knowledge Graphs. We can use the Graph Index and Other representations to tell the Taxonomies, for lets say, a R. System **Knowledge Graph**.

Knowledge graphs can be made by using XLSX/JSON for better SEO to leveraging taxonomy ie. Combining knowledge graphs and machine learning technologies.

We need to explore the naturally occurring structure of **VIL's metadata** to come up with an appropriate ontological structure (e.g. creating communities). Using **CUSTOMER ENGAGEMENT VARIABLES** we should be able to identify patterns in the data.

Used cases [VIL's PS specific]

> - Develop a new Legacy system to store taxonomies of Customer Behaviour and preferences and give suggestions of a product to customers

> - Recommend benefits or rewards that will help VIL and partners nudge customer behavior toward business intents: product and services.

> - Online Directory of [VODAFONE IDEA LIMITED]

Graph databases understand the connections in large and varied data sets Ease Visualisation and even help Finding Trilogies Analysis of graph structured data help us to provide invaluable insights about the relationships between entities, as well as enable optimisations over a network of interconnected objects [w.r.t. Conventional Methods] , Much closer to our whiteboard model Calculate best route by certain algorithms [DIJKSHTRA's and other graph Algo] of connected semi structured data.

HIGH LEVEL SOLUTION ARCHITECTURE

Data Transformation

Here we need to build a process that consumes several input **Customer engagement variables** and creates a collated csv with the relevant features

Data Modelling

We need to build a weighted undirected graph. We should explore the different pros and cons of a sparse or dense graph . Labelled entities in the nodes. The structure should be built in such a way it can help in addressing the set of identified use-cases Other ways to build the ontology other than using just occurrences.

Data Operations

Be able to save the data to a database or in an appropriate file-format, add and delete data from it

Finally we should be able to use it for our use-case by extracting information from it.

TECHNOLOGY STACK
VIL CODEFEST IDEA SUBMISSION

For Development, **JavaScript and Flask** will be used for building the Solution. **Jupyter Notebook** will be used as the IDE for coding the idea in Python for the machine ,earning models implemented using **Keras and Tensorflow.**

ML STACK

With Sklearn, Keras and Tensorflow, to run on top of CNTK, TensorFlow, or Theano. It's ideal for projects that demand a deep learning library that accommodates rapid prototyping through modularity and extensibility. It also supports recurrent and convolutional networks, and seamlessly runs on both CPUs and GPUs.

Github will offer a platform to open source and host the code for reusability and displaying to the Tech team.

✦**PROGRAMMING LANGUAGES -** Python, C++ AND JS.
✦**FRAMEWORKS & LIBRARIES -** Keras, TF AND NodeJS
✦**SERVERS -** Local and GCP.
✦**UI/UX solutions -** ADOBE, Bootstrap and Flask.
✦**Other tools TO BE used -** Github, Stack-Overflow AND Jupyter Notebook

We should identify all possible use-cases (in addition to the ones mentioned above) and come up with an appropriate structure for our content based taxonomy. Once we have identified the structure, we should implement a prototype of an end to end pipeline which can do the following

- Ingest raw content and process it to create the relevant input data for creating a Graph Structure.

- Create a taxonomy out of it using some API / Platform / Library / Open source framework / Architecture.

- Save the data to a database or in an appropriate file-format Add data to that taxonomy (structure should update accordingly) to make it more robust across shows / genres

- Delete data from the taxonomy (structure should update accordingly)

Perform Common Data Operations like Visualising it for the different use-cases and extracting information which can help in addressing the different use-cases.