Content Based Recommender System using Unsupervised Learning

Executive summary report for the Sales Marketing Team Prepared by Data Science

ISSUE / PROBLEM

The challenge we encountered was defining a value to assign to the more engaging and relevant courses to learners. We also considered ethical consequences of building this model, particularly regarding the errors that we will encounter when the model is built.

RESPONSE

What's likely to give the best recommender system?

We built three different models **User Profile** and **Course Genres**, **Course Similarity** and **Clustering** compared results and parameters.
We used expertise, If the dataset contains detailed course attributes and user preferences, clustering-based approaches may provide more personalized recommendations. However, if the dataset is sparse or lacks detailed attributes, a course similarity approach might be more suitable.

IMPACT

This model helps predict courses that best suit the user profile and course profile. This recommendation can assist users in making decisions about their next courses, ultimately increasing engagement and revenue streams.

	user	item	rating	COURSE_ID	TITLE
0	733707	RP0103	3.0	RP0103	using r with databases
1	733707	BD0212EN	3.0	BD0212EN	spark fundamentals ii
2	733707	BD0211EN	3.0	BD0211EN	spark fundamentals i
3	733707	ST0101EN	3.0	ST0101EN	statistics 101
4	733707	BD0115EN	3.0	BD0115EN	mapreduce and yarn
18	733707	DV0101EN	3.0	DV0101EN	data visualization with python
19	733707	SC0105EN	3.0	SC0105EN	data science with scala
20	733707	BD0145EN	3.0	BD0145EN	sql access for hadoop
21	733707	DB0151EN	3.0	DB0151EN	nosql and dbaas 101
22	733707	BD0131EN	3.0	BD0131EN	moving data into hadoop

Table above shows the recommended courses for user 733707 using user profile and course genre

	USER	COURSE_ID	SCORE	TITLE
3870	733707	ML0120ENv3	1.000000	deep learning with tensorflow
3871	733707	ML0120ENv2	1.000000	deep learning with tensorflow
3872	733707	ML0122ENv1	0.982873	accelerating deep learning with gpu
3873	733707	DS0110EN	0.732941	data science with open data
3874	733707	excourse67	0.708214	introduction to big data
3875	733707	excourse72	0.703648	foundations for big data analysis with sql
3876	733707	excourse63	0.694563	a crash course in data science
3877	733707	excourse46	0.689253	machine learning
3878	733707	excourse47	0.680065	machine learning for all
3879	733707	ML0101ENv3	0.662622	machine learning with python

Table above shows the recommended courses for user 733707 using course similarity

	user	cluster	rec_1	rec_2	rec_3	title_1	title_2	title_3
102	1078030	1	PY0101EN	DA0101EN	DV0101EN	python for data science	data analysis with python	data visualization with python
151	674939	28	BD0111EN	BD0115EN	BD0211EN	hadoop 101	mapreduce and yarn	spark fundamentals i
221	2057052	16	DS0101EN	BD0101EN	PY0101EN	introduction to data science	big data 101	python for data science
298	733707	9	BD0111EN	PY0101EN	BD0211EN	hadoop 101	python for data science	spark fundamentals i

In the Clustering model above shows the recommendation results

INSIGHTS/NEXT STEPS

- **If you're exploring the structure or patterns within the data without labeled examples, unsupervised learning is more appropriate. It can help in understanding the underlying structure of the data and finding hidden patterns.**
- **Unsupervised learning algorithms like k-means clustering can be useful for segmenting data into distinct groups based on similarities.**
- **Techniques like Principal Component Analysis (PCA) are used for dimensionality reduction and visualization, which can be valuable for understanding high-dimensional data.**
- **Unsupervised learning is often used for anomaly detection where the goal is to identify rare events or outliers in the data.**
- **Considering human suggestion (their level) and Hybrod models.**