User-Specific Cryptographic Asset Recommendation System

By: Luis, Dawn,Kevinray, Connor

Business Case & Project Description

Business Problem

RMDS wants to increase usage of NFT & user engagement on platform

Solution

Integrated NFT Recommendation System

- Legitimacy
- Liquidity
- Engagement

Project Description

Created & Designed NFT Recommendation System

- Created a Supervised Machine Learning model (SVM) & Natural Language Processor (NPL)
- Built a User similarity Score to provide NFT recommendations
- Redesigned RMDS Platform to drive engagement

Datasets

NFT classification

- NFT historical sales dataset
- 108147 rows of NFT data
- 7 selected feature set

```
In [12]: df2.head()
```

Out[12]:

	collection_name	asset_id	asset_name	asset_description	event_quantity	event_payment_symbol	event_total_price
0	Rarible	18214580	Daft Punk Never Die	Piece art Daft Punk always hears Size x	1.0	ETH	0.070000
1	Rarebit Bunnies	18276844	Rarebit #164 - Wax Off Bunny	When comes high kicks Rarebit back paws lethal...	1.0	ETH	0.150000
2	Rarible	16911700	Meditation	Meditation Diana Minted NFT collectibles	1.0	ETH	0.001000
3	Rarible	16986936	I´m OG	This one first NFTs human history	1.0	ETH	0.000647
4	ChainGuardians	13382164	Celia B100 #105	One original androids created within Chainguar...	1.0	ETH	0.200000

User Similarity

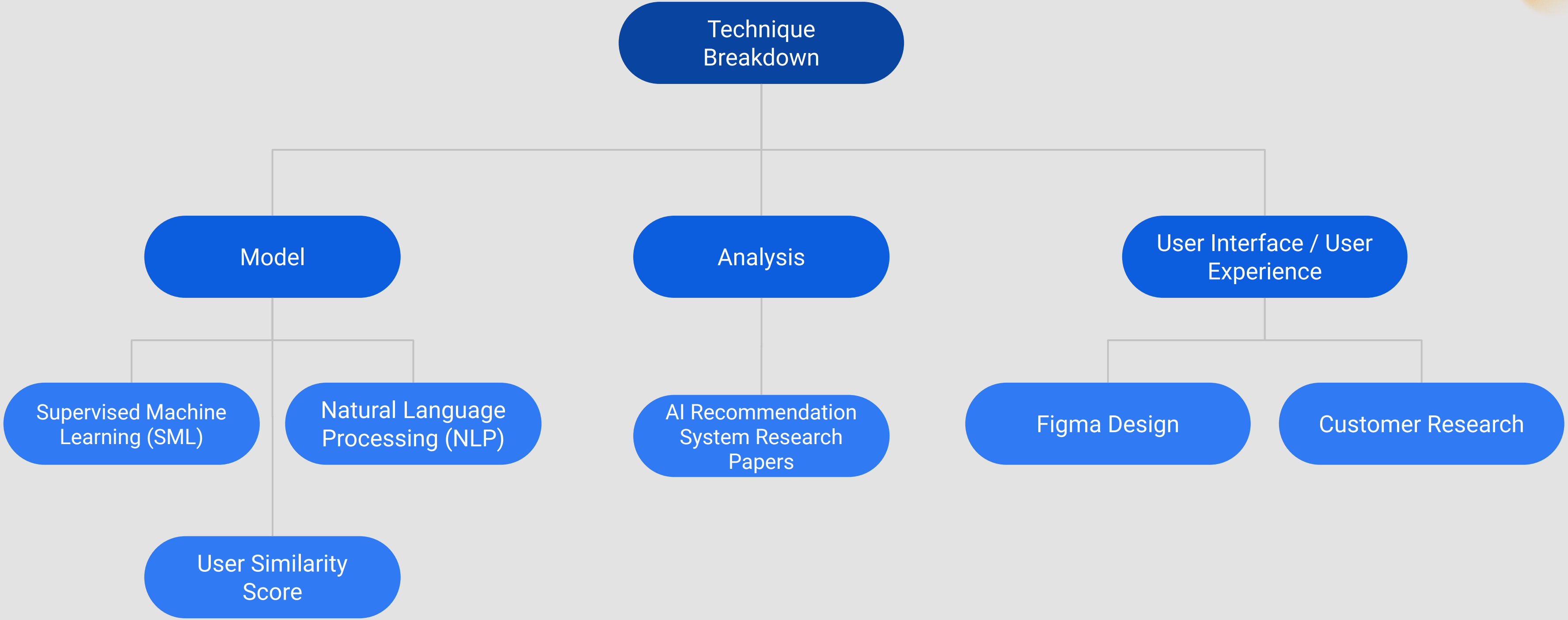
- Stack Overflow dataset
- 50000 rows of user data
- 14 selected feature set

```
In [7]: df.head()
```

Out[7]:

MainBranch	Employment	Country	EdLevel	DevType	OrgSize	Currency	CompTotal	CompFreq	Age	Gender	Ethnicity	Accessibility	MentalHe
I am a developer by profession	Independent contractor, freelancer, or self-em...	Slovakia	Secondary school (e.g. American high school, G...	Developer, mobile	20 to 99 employees	EUR European Euro	4800.0	Monthly	25-34 years old	Man	White or of European descent	None of the above	None of ab
I am a student who is learning to code	Student, full-time	Netherlands	Bachelor's degree (B.A., B.S., B.Eng., etc.)	NaN	NaN	NaN	NaN	NaN	18-24 years old	Man	White or of European descent	None of the above	None of ab
I am not primarily a developer, but I write co...	Student, full-time	Russian Federation	Bachelor's degree (B.A., B.S., B.Eng., etc.)	NaN	NaN	NaN	NaN	NaN	18-24 years old	Man	Prefer not to say	None of the above	None of ab
I am a developer by profession	Employed full-time	Austria	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	Developer, front-end	100 to 499 employees	EUR European Euro	NaN	Monthly	35-44 years old	Man	White or of European descent	I am deaf / hard of hearing	N
I am a developer by profession	Independent contractor, freelancer, or self-em...	United Kingdom of Great Britain and Northern I...	Master's degree (M.A., M.S., M.Eng., MBA, etc.)	Developer, desktop or enterprise applications;...	Just me - I am a freelancer, sole proprietor, ...	GBP\tPound sterling	NaN	NaN	25-34 years old	Man	White or of European descent	None of the above	N

Techniques Breakdown



Results Table - Grouping

Model

Validation results using 70,000 Training cases & 30,000 Testing Cases:

SVM 1:

```
Accuracy (RBF Kernel):  87.33
F1 (RBF Kernel):  85.01
```

SVM 2:

```
Accuracy (RBF Kernel):  95.59
F1 (RBF Kernel):  94.60
```

Final:

```
Accuracy (RBF Kernel):  97.10
F1 (RBF Kernel):  96.58
```

K-Means (K = 20)

SSE 6982.01

Silhouette Score: 0.5815

Baseline

Accuracy Level: 97% for SVM (Song et al, 2018)

K-Means:

1. Standardization needs to be and can affect the performance of K-Means. Also, datasets have different scales of measurement (Su et al, 2009)
2. Typical 2-4 clusters depending on datasets but can increase based on the data size to improve the model (Pham & Nguyen, 2005)

Results Table - Similarity Score

Model

Validated on 50,000 users dataset

Top 3 recommendation for a user most similar to someone from USA

```
In [180]: recommendation('United States of America')

          (20836, 0.8577661764387349) United States of America
          (28177, 0.8577661764387349) United States of America
          (36353, 0.8520209591874592) United States of America
```

Sample output, to find the user similarity in cases of cold start for a new user.

Baseline

1. TF-IDF & Cosine similarity matrix

2. "CountVectorizer method creates a highly sparse representation of a document over the vocabulary". (Shahmirzadi *et al 2018*)

3. Cosine similarity has higher positive correlation in comparison to Euclidean (Adeyanju *et al, 2018*)

Results table - Final Framework

Framework

1. Combined both previous models (SVM & Similarity Scores)
    a. Generates a User Specific Recommendations
2. Incorporates user clicking history dataset

3	10248	[]
4	2058	[https://opensea.io/assets/0x629a673a8242c2ac4...
...	...	...
49995	11889	[https://opensea.io/assets/0x495f947276749ce64...

Implementation

1. Divided users into two groups
2. Gathers data from clicking history get URLs of NFTs users visited
3. Acquires needed features from URLs, then fit features to predict preferred collections
4. Cold Start → Recommends similar users' preferences

```
array(['sorare', 'sorare', 'tekhne', 'beeple', 'rarible'], dtype=object),
```

5. Incorporated price filter and trending products into model

Demo: Final Outputs

Example case

User ID: 2052

User type: new

```
Test case for user 2052 : new user
```

NFT Group Prediction (based on clicking histories of similar users)

```
[array(['cryptokitties', 'sorare', 'crypto-geishas', 'sorare', 'rarible',
       'cryptopunks', 'cryptokitties', 'cryptokitties', 'superrare',
       'rarible', 'sorare', 'sorare', 'sorare', 'rarible', 'superrare'],
      dtype=object), array(['sorare', 'sorare', 'sorare', 'rarible', 'sorare', 'rarible',
       'axie', 'the-sandbox-assets', 'axie', 'sorare', 'sorare', 'sorare',
       'sorare', 'cmyk', 'sorare', 'rarible', 'axie', 'sorare',
       'hashmasks', 'twerky-bags', 'sorare', 'rarible', 'rarible',
       'rarible', 'sorare', 'ether-horizon', 'pills', 'sorare',
       'hashmasks', 'sorare', 'sorare', 'sorare', 'sorare', 'ethermon',
       'sukamii', 'hashmasks', 'nightfamily', 'decentraland', 'sorare'],
```

Price filtering in ETH

```
Enter a lower bound for price in ETH:  0
Enter a higher bound for price in ETH:  1000
```

Recommendation

```
recommending user 2052 [('axie', 3), ('hashmasks', 6), ('sorare', 30)]
```

Demo - UI (User Interface)

UI RECORDED DEMO: Click Here

UI LIVE DEMO: Click Here

Deliverables Explanation

1. **Team Skill Sets & Background:** 2 coders and 2 designers
2. **Goal:** Maximize personal utility, reflect personal skill sets
3. **Progress:** 97.1% accuracy on backend model, comprehensive/point on implementation representing algorithm/suggestions to stakeholder.
4. **Concern:** Diminishing return for code implementation. Contribution of designers goes down, focus shift/workload inflation for coders.
5. **Claim:** Non-coding UI prototype is NOT a representation of slacking off, but an unanimous decision within the group from the start of the semester to deliver a better project & product to the stakeholder

Improvements

UI & Backend Connection

Continuous Improvement of Accuracy

Increase Customer Research

References

1. Beel, J., Langer, S., Genzmehr, M., Gipp, B., Breltinger, C., & Numberger, A. (2013). Research Paper Recommender System Evaluation: A Quantitative Literature Survey. *International Journal on Digital Libraries*, 1-34. doi: 10.1007/s00799-015-0156-0
2. Garcia, S., Gallego-Ramirez, S., Luengo, J., Benitez, J., & Herrera F. (2016). Big data preprocessing: methods and prospects. *Big Data Analytics*, 1,9. doi: https://doi.org/10.1186/s41044-016-0014-0
3. Ioffe, S., & Szegedy, C. (2015). *Batch normalization: Accelerating deep network training by reducing internal covariate shift*. Cornell University Library, arXiv.org.
4. Kozak, S. (2018). Shrinking the Cross Section. Retrieved from, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2945663
5. Pham, Dimov, S. S., & Nguyen, C. D. (2005). Selection of K in K-means clustering. Proceedings of the Institution of Mechanical Engineers. Part C, Journal of Mechanical Engineering Science, 219(1), 103–119. https://doi.org/10.1243/095440605X8298
6. Putri, D., Leu, J., Seda, P. (2019). Design of an Unsupervised Machine Learning-Based Movie Recommender System. *Symmetry*, 12,185. doi: doi:10.3390/sym12020185
7. Piyadigama, D., & Poravi, G. (2022). Exploration of the possibility of infusing Social Media Trends into generating NFT Recommendations. *Cornell University RecSys*. doi: https://doi.org/10.48550/arXiv.2205.11229
8. Piyadigama, D., & Poravi, G. (2022). An Analysis of the Features Considerable for NFT Recommendations. *Cornell University*. doi: https://doi.org/10.48550/arXiv.2205.00456
9. Song, D., , Y., Min, Q., Sun, Q., Ye, K., Zhou, C., Yuan, S., Sun, Z., Liao, J. (2018). Similarity-based machine learning support vector machine predictor of drug-drug interactions with improved accuracies. *Journal of Clinical Pharmacy and Therapeutics*, 44, 268-275. doi: https://doi-org.libproxy1.usc.edu/10.1111/jcpt.12786
10. Su, C. , Zhan, J., & Sakurai, K. (2009). Importance of Data Standardization in Privacy-Preserving K-Means Clustering. Database Systems for Advanced Applications, 5667, 276–286. https://doi.org/10.1007/978-3-642-04205-8_23
11. Syakur, M., Khotimah, B., Rochman, E., Satoto, B. (2017). Integration K-Mean Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster. *Materials Science and Engineering*, 336. doi: 10.1088/1757-899X/336/1/012017