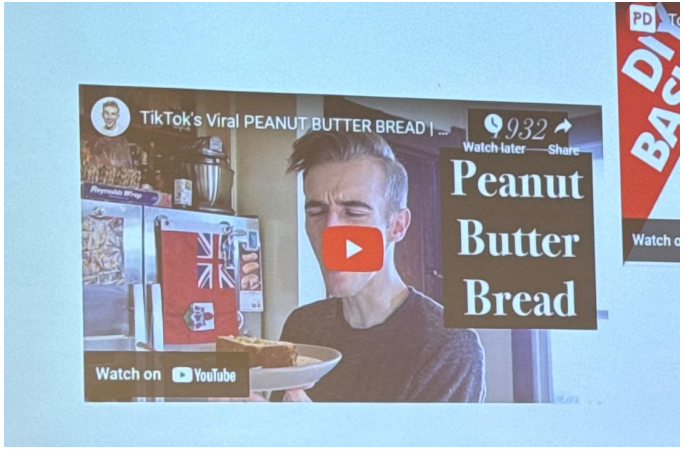
TrendCart

Integrating Viral Food Trend Predictions into Grocery Stores Demand Forecasting

Sapan Patel, Akaash Dash

Motivation

- Social media platforms have grown rapidly in popularity
- Viral trends can spread like wildfire.
- Grocery stores can take advantage of these trends

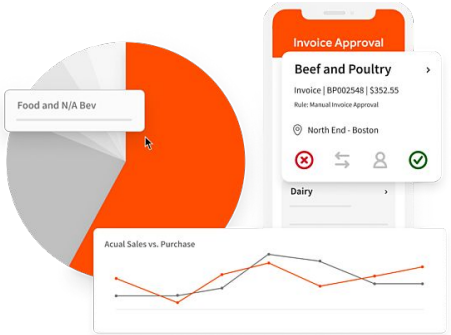


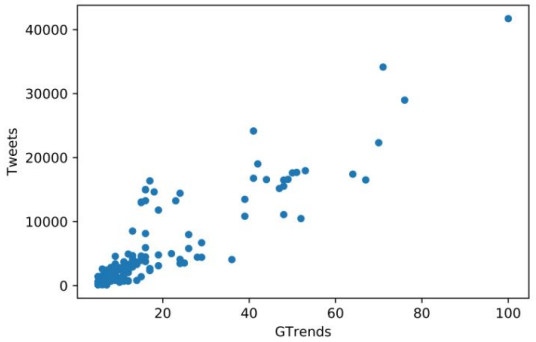
Example of Viral TikTok Recipe video

Recipe book containing Viral TikTok Recipes

Related Works

- Platforms: MarkT POS and Toast Inventory Management



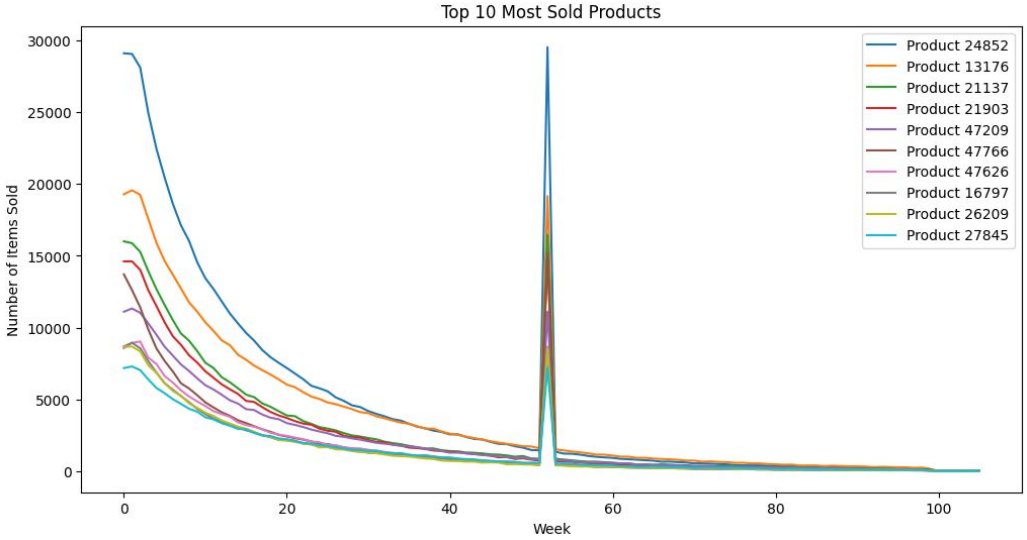


- Academic Works:
  - Correlation between google trends and twitter engagements
  - Predicting movie box office sales by surveying google trends

Dataset - Instacart Basket

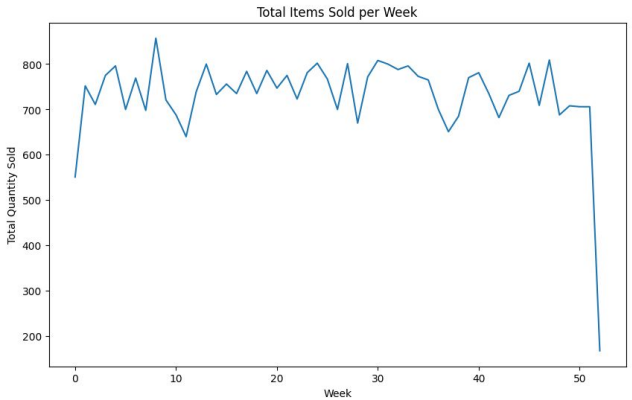
- Was not a good fit for our project
  - Was intended for exploring products bought together and repeated purchases
- Too granular
- Not timestamped

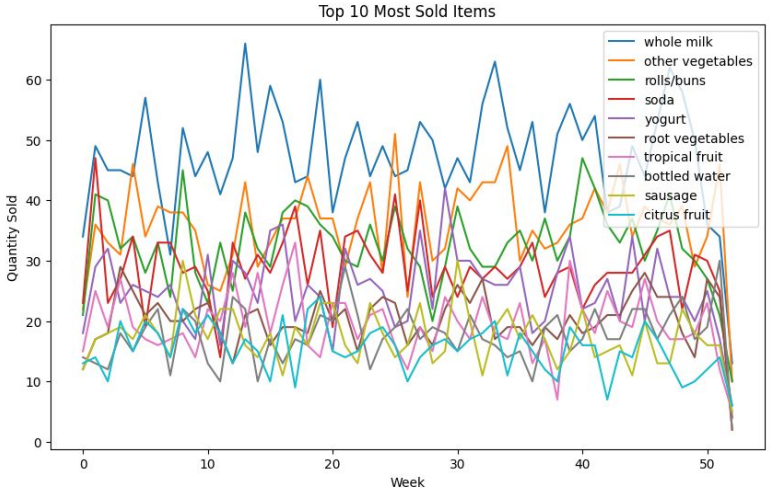
```
49678,Large Chicken & Cheese Taquitos
49679,Famous Chocolate Wafers
49680,All Natural Creamy Caesar Dressing
49681,Spaghetti with Meatballs and Sauce Meal
49682,California Limeade
49683,Cucumber Kirby
49684,"Vodka, Triple Distilled, Twist of Vanilla"
49685,En Croute Roast Hazelnut Cranberry
49686,Artisan Baguette
49687,Smartblend Healthy Metabolism Dry Cat Food
49688,Fresh Foaming Cleanser
```



Dataset - Groceries

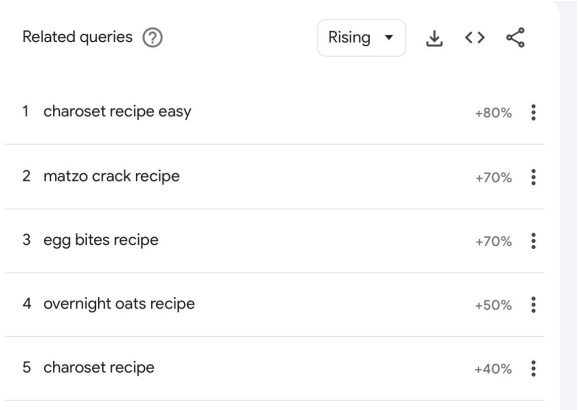
- Consisted of sales at a small grocery store
  - Provided purchase date and product for each purchase
- First step was to create a library of product IDs
  - Observed that some are categories of products not products
- Next, created a timeseries of data
  - Aggregated sales by product/category per week





Dataset - Trends

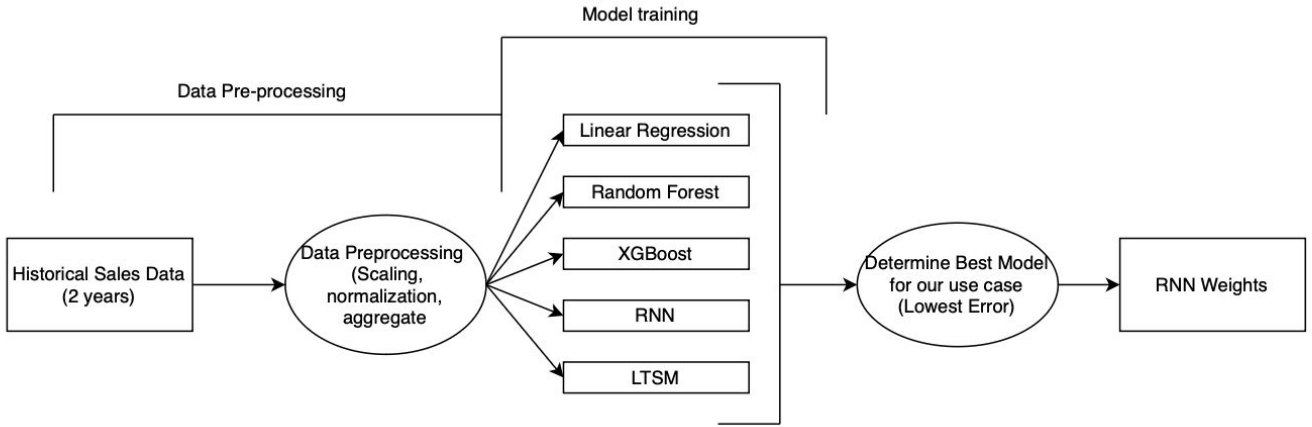
- No social media APIs accessible for individual trend data
- Google Trends to detect rising food trends/viral recipes
- Historical dataset unobtainable
  - Have to query by week, but API is heavily rate limited
- Can only use trends at inference time per week



Dataset - Recipe Ingredients

- Spoonacular API
  - Allows for searching their recipe database as well as scraping ingredients from websites

Architecture - Underlying ML Model Training



1. Pre-process historical data
   a. Scale, aggregate, remove irrelevant data points
   b. Normalize with MinMaxScaler from Scikit-Learn
2. Train different ML models and compare performance
   a. Evaluate the best fit through comparing the loss function values
3. Save the weights for the best model

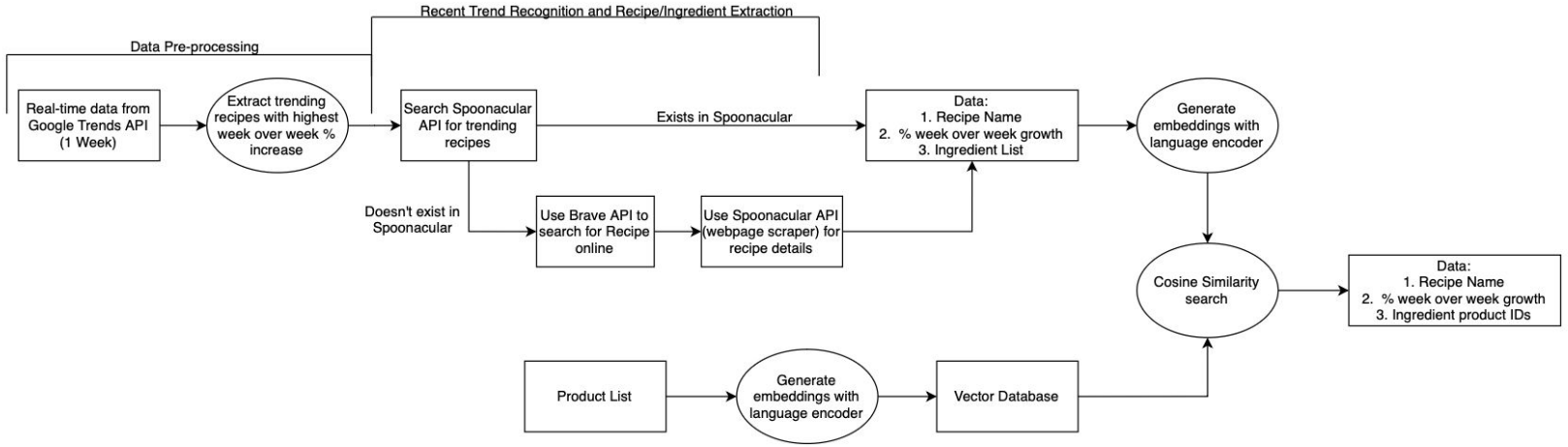
Results - Underlying ML Model Training

```
Linear Regression R^2 Score: -0.9981017506615912
Random Forest R^2 Score: -0.15533955861504445
XGBoost R^2 Score: -0.6429575482844548
```

```
LSTM Mean Squared Error: 0.05794665217399597
```

```
RNN Mean Squared Error: 0.06031983718276024
```

Architecture - Trending Recipe Scraper



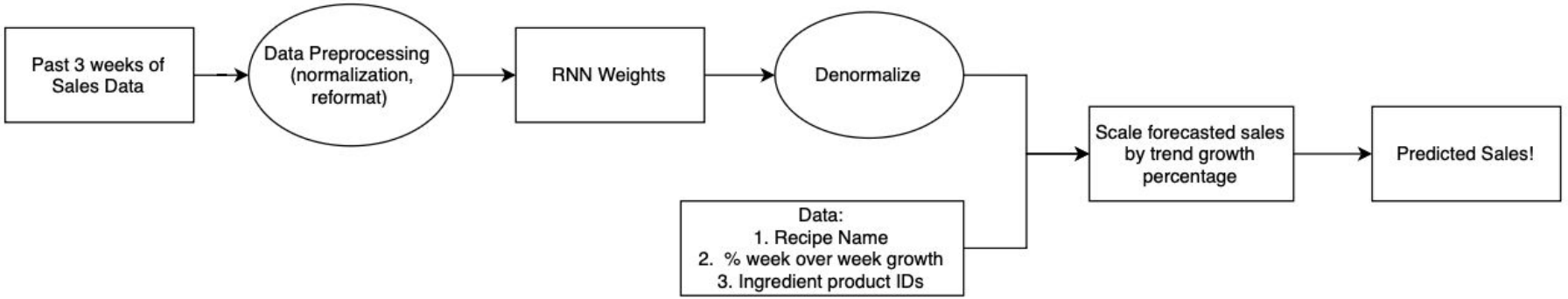
1. Query Google Trends API for recipes
2. Search Spoonacular database for recipe and get ingredients
3. If not in Spoonacular database, scrape from web
    a. Search recipe online using Brave Search API
    b. Query Spoonacular API with link to scrape site and gather ingredients
4. Use language encoder model and cosine similarity search to label ingredients

Results - Recipe Trend Scraper


```
{
    "name": "pimento cheese recipe",
    "growth": 200,
    "ingredients": [
        "cheddar cheese",
        "cream cheese",
        "pimento peppers**",
        "mayonnaise",
        "garlic powder",
        "onion powder",
        "ground cayenne pepper",
        "jalape\u00f1o pepper",
        "pepper",
        "salt"
    ]
},
{
    "name": "hummingbird food recipe",
    "growth": 50,
    "ingredients": [
        "carrot",
        "baking soda",
        "spices",
        "all purpose flour",
        "salt",
        "olive oil",
        "vanilla",
```

```
{
    "recipe": "pimento cheese recipe",
    "growth": 200,
    "ingredients": [
        129,
        66,
        163,
        72,
        113,
        60
    ]
},
{
    "recipe": "hummingbird food recipe",
    "growth": 50,
    "ingredients": [
        1,
        34,
        35,
        4,
        69,
```

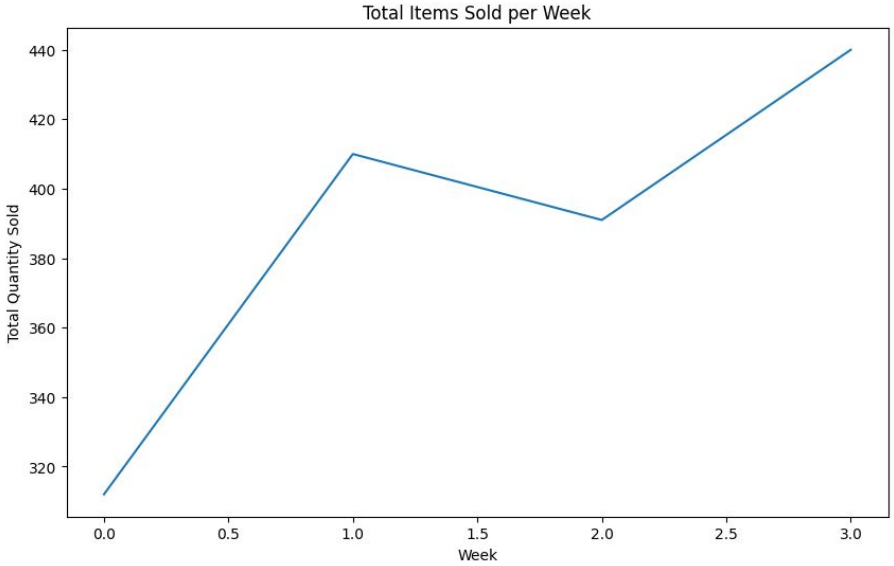
Architecture - Combined Predictive Analysis

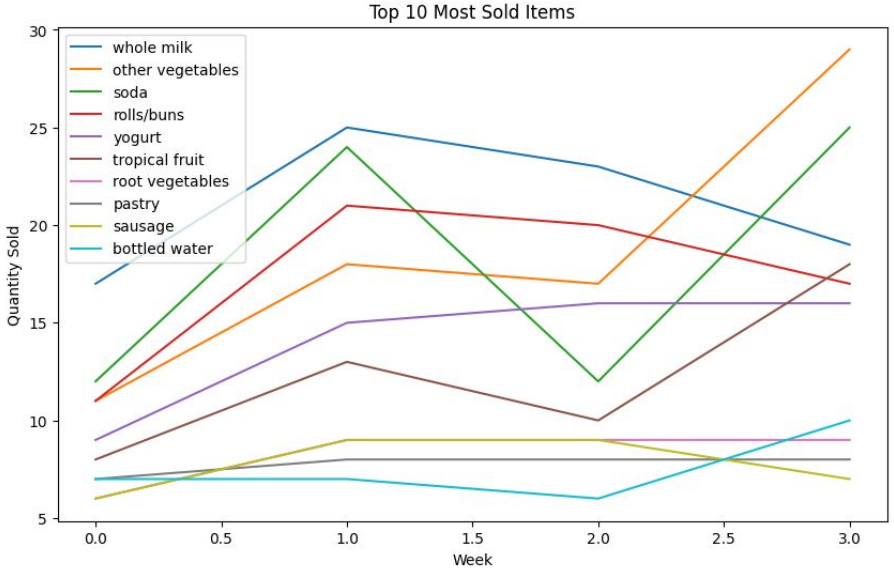


Used at forecasting time:

1. Preprocess real-time past three weeks sales data
   a. Reformat and normalize using MinMaxScaler from Scikit-Learn
2. Pass data to saved RNN model weights for inference, obtain initial prediction
3. Denormalize initial prediction to match original scale
4. Augment ingredients in viral/trending recipes by week-over-week growth

Results - Combined Predictive Analysis





Implementation Details

- TrendCart was fully implemented using Python
- Past sales data is contained in csv format, but converted to npy files
- Data from API responses are in JSON format
- Pre-processing was done using the Pandas library
- Since we are training many different models
  - SciKit-Learn: Linear Regression and Random Forest Models
  - TensorFlow: LSTM and RNN
  - XGBoost: XGBoost Model
- Finally, we used MatPlotLib for the graphs and visuals since it is the most extensive library for that use

Limitations

- We assume that Google Trends captures true virality, but some might spread through other mediums and not captured in Google Trends
    - Additionally, Google Trends data is aggregated and may not capture localized spikes

- We assume that Google Trends virality translates to real world sales: a spike in Google Trends data might show interest in a topic, it does not guarantee sales (cyber-physical gap)

- We assume static conditions and no black swan events such as the SouthWest Airlines Incident, limiting TrendCart to ‘Virtual Time’

- Fourth, based on proposal feedback, we focused on Trend Recognition and Predictive analysis, resulting in a weak frontend

- We predict the ML models will struggle with knowledge obsolescence. Long term shifts in consumer preferences may risk the forecasting abilities of TrendCart

Skills Learned

- We gained valuable experience in the cyber-physical gap.
  - The entire premise of TrendCart pertains to this as we are aiming to bridge the gap between cyber trends and physical sales.

- Trend recognition is another hugely important skill, and implementing this project allowed us to gain exposure to what trends look like, how to recognize them, and how they can be applied/used to provide value.

- Finally, another skill we gained exposure to is Situational Awareness.
  - This primarily comes in the form of identifying assumptions. When starting this project, we had many assumptions of what data and technologies exist and how things work, but these were quickly proven wrong.

Technical Skills

- The implementation of the Grocery Sales Forecasting and Inventory Management Dashboard encompasses a comprehensive skill development plan for us.
  - Dataset cleaning and integration
  - API Integration (pytrends, spoonacular, Google Trends API, brave API)
  - Web Scraping techniques (Spoonacular, Beautiful Soup)
  - Data analysis and machine learning
  - Feature engineering and data processing
  - Hyperparameter tuning

Changes from Proposal

- We were given a lot of feedback to have more realistic expectations and focus on the core product of viral trend recognition:
  - Dropped the website
  - Anyway inventory restock would just be a matter of checking if demand > restock
- Therefore, we focussed our efforts on 2 things based on feedback:
  - Real-Time trend detection and Sales prediction as these are the more innovative challenges
  - They are very relevant to a few key issues: bridging the cyber-physical gap through trend prediction and utilizing machine learning for sales prediction based on “pseudo” real-time data

References

[1] Accenture. “Why the Future of Shopping Is Set for a Social Revolution.” *Accenture.com*, Accenture, 2 Jan. 2022, www.accenture.com/us-en/insights/software-platforms/why-shopping-set-social-revolution.

[2] Hamstra, Mark. “Sure, TikTok Is Fun. But Can It Drive Sales?” *Supermarket News*, 14 Dec. 2022, www.supermarketnews.com/online-retail/sure-tiktok-fun-can-it-drive-sales.

[3] Evans, Michelle. “How Gen Z Inspired Grocers to Reinvent Their Social Presence.” *Forbes*, 29 Nov. 2022, www.forbes.com/sites/michelleevans1/2022/11/29/how-gen-z-inspired-grocers-to-reinvent-their-social-presence/?sh=639ba5484692. Accessed 10 Feb. 2024.

[4] Youn, Soo. “Trader Joe’s Sold out of Kimbap, Its Latest Viral Offering, Thanks to TikTok.” *NBC News*, 6 Sept. 2023, www.nbcnews.com/news/asian-america/trader-joes-kimbap-tiktok-rcna103403.

[5] Kalatzis, Nikos & Roussaki, Ioanna & Matsoukas, Christos & Paraskevopoulos, Marios & Papavassiliou, Symeon & Tonoli, Simona. (2018). Social Media and Google Trends in Support of Audience Analytics: Methodology and Architecture.

[6] Schoen, Harald & Gayo-Avello, Daniel & Metaxas, Panagiotis & Mustafaraj, Eni & Strohmaier, Markus & Gloor, Peter. (2013). The power of prediction with social media. Internet Research: Electronic Networking Applications and Policy. 23. 10.1108/IntR-06-2013-0115.

[7] Google. (n.d.). *Google Trends*. Google trends. https://trends.google.com/trends/explore?geo=US&q=recipe&hl=en-US

[8] Urbansky, D. (n.d.). *Documentation*. spoonacular recipe and food API. https://spoonacular.com/food-api/docs#Extract-Recipe-from-Website

[9] *Instacart Market Basket Analysis*. Kaggle. (n.d.). https://kaggle.com/competitions/instacart-market-basket-analysis

[10] Dedhia, H. (2020, September 17). *Groceries dataset*. Kaggle. https://www.kaggle.com/datasets/heeraldedhia/groceries-dataset

[11] *Markt POS | Home*. (n.d.). Www.marktpos.com. Retrieved April 21, 2024, from https://www.marktpos.com/

[12] Toast. (n.d.). *Toast | Restaurant Point of Sale & Management System*. Pos.toasttab.com. https://pos.toasttab.com/

[13] Shahbaznezhad, H., Dolan, R., & Rashidirad, M. (2021). The Role of Social Media Content Format and Platform in Users’ Engagement Behavior. Journal of Interactive Marketing, 53(1), 47-65. https://doi.org/10.1016/j.intmar.2020.05.001

[14] *Sklearn.preprocessing.MinMaxScaler*. scikit. (n.d.). https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html

[15] Brave Search API. (n.d.). Brave. Retrieved April 21, 2024, from https://brave.com/search/api/

[16] Sitikhu, P., Pahi, K., Thapa, P., & Shakya, S. (2019, October 30). *A Comparison of Semantic Similarity Methods for Maximum Human Interpretability*. ArXiv.org. https://doi.org/10.48550/arXiv.1910.09129

[17] Kennedy, J. (2024, March 15). *Who Is the TikTok Shopper?* The Business of Fashion. https://www.businessoffashion.com/articles/retail/who-is-the-tiktok-shopper/

[18] Pizza, H. R. I., & Mary. (2023, August 16). *Culinary Confessions: Cooking Habits of Gen Z & Millennials*. Home Run Inn Pizza. https://www.homeruninnpizza.com/news-blog/fun-facts/cooking-habits-gen-z-millennials/