International Conference on
Machine Intelligence and Emerging Technologies
September 23-25, 2022
Noakhali Science and Technology University (NSTU), Noakhali, Bangladesh
https://confmiet.org/

Assorted, Archetypal and Annotated Two Million (3A2M) Cooking Recipes Dataset based on Active Learning

Nazmus Sakib , G. M. Shahariar , Md. Mohsinul Kabir, Md. Kamrul, Hasan Mahmud

Presented By:

Nazmus Sakib

Paper id: 462 | September 23, 2022

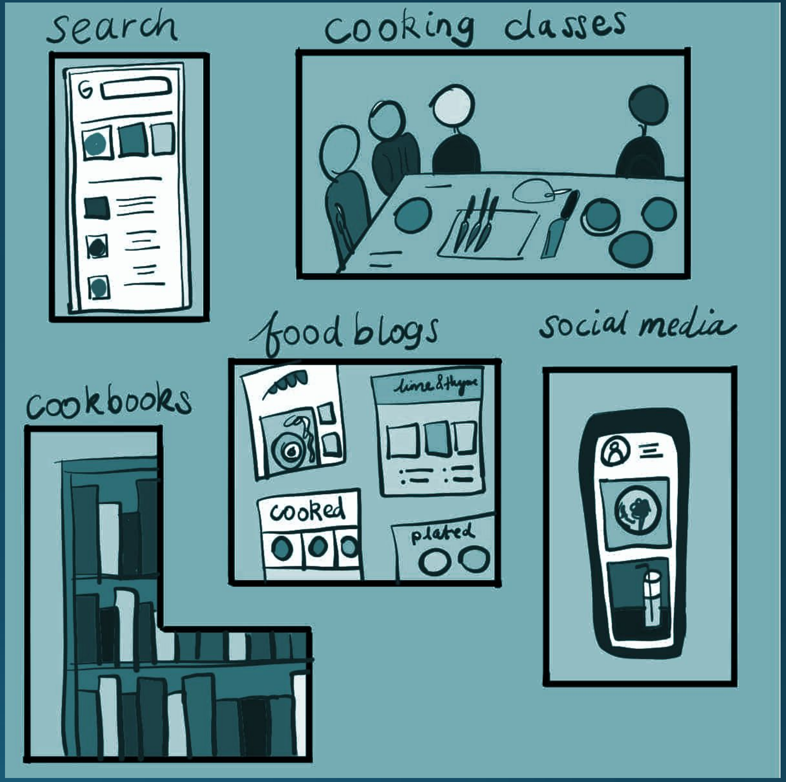
Springer EAI RESEARCH MEETS INNOVATION

IEEE COMPUTER SOCIETY Bangladesh Chapter

http://CenNSER.org CNSER Center for Natural Science & Engineering Research

Introduction

- Recipes are used since many decades as a guideline for cooking, but same recipe might be represented in different ways.
- Recipes divided into genres can allow customers to make an informed decision based on their interest.
- Considering the abundance of data, culinary recipe datasets are attracting a lot of attention recently.
- There is a lack of properly annotated culinary dataset based on the judgement of domain experts.



Motivation

- **Easier Food Decisions** — People can easily bake their desired genres of food following the recipe with their available items.
- **Dataset For Research** — Utilizing this massive dataset can pioneer new areas of research like recipe generation, genre classification etc.
- **Medical Research** — Medical sectors, particularly those working with food nutrition, might advise food variations based on these datasets.

Literature Review

M. Bien, M. Gilski, M. Maciejewska, W. Taisner, D. Wisniewski, and L. A., “Recipenlg: A cooking recipes dataset for semi-structured text generation,” in InProceedings of the 13th International Conference on Natural Language Generation, Dec. 2020, pp. 22–28.

- The RecipeNLG collection is the largest accessible dataset in the domain, with 2,231,142 different culinary recipes from various sources such as cookbooks, blogs, and recipe websites.
- A list of substances known as Name Entity Recognition (NER) is presented in the dataset, but same items may occur in many recipes.
- One of the shortcomings of this dataset is that the genre of the recipes were discovered to be uncategorized or unclassified.

J. Marin, A. Biswas, F. Ofli, et al., “Recipe1m+: A dataset for learning cross-modal embeddings for cooking recipes and food images,” IEEE transactions on pattern analysis and machine intelligence, Jul, vol. 43, no. 1, pp. 187–203, Jul. 2019.

- Recipe1M+ is a large-scale, organized corpus containing over one million culinary recipes and 13 million food photos.
- Authors trained a neural network to learn a combined embedding of recipes and photos, which produced outstanding results on an image-recipe retrieval test.
- Utilizing this dataset, a method was developed for autonomously producing culinary recipes. Translation metrics were used to evaluate the model.

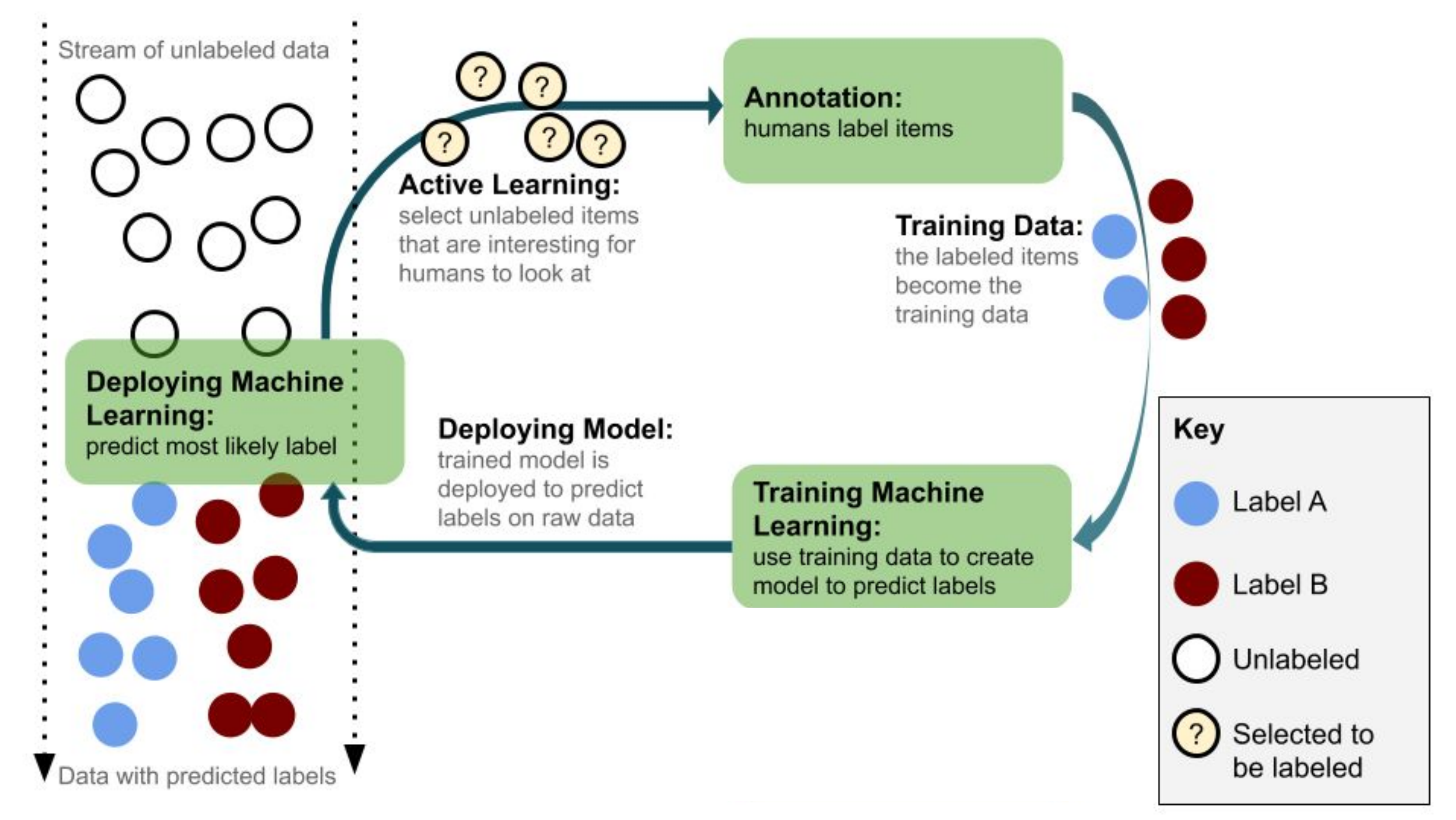
B. Settles, "Active learning literature survey," 2009.

- Active learning is a supervised machine learning technique that trains a predictor iteratively and utilizes the predictor to pick the training instances in each iteration, boosting the predictor's odds of selecting better configurations and improving the prediction model's accuracy.

- Active learning is an effective method to label the unlabeled data.

- Main challenge of active learning- computational cost.

Labeling Data Through Active Learning



[Image Reference: active learning tensorflow]

Challenges in Culinary Recipes Research

- Categorizing raw recipes to the appropriate food genres is a challenging task in this domain as there are many conflicting dish names, cooking techniques, ingredients, and recipe sources.
- Utilizing the knowledge of domain experts to categorize recipes could be a solution.
- Human in the Loop is required for this form of categorization since it is heavily reliant on the Human annotator. However, for large datasets such as 2 million, the ensemble approach is useful.

Research Question

- Can active learning using human-in-the-loop help categorize any recipe based on name, direction and the ingredients of the food from online resources?

Objectives with Specific Aims

- To construct a recipe dataset that contains nine genres developed by domain experts.

- To apply active learning and ensemble-based techniques to semi-automate the annotation process of 2 million data using human-in-the-loop approach.

Possible Outcomes

- An annotated original dataset of two million culinary recipes that developed by the judgement of domain experts.

- A novel approach to annotate culinary recipes through active learning and ensemble techniques, with a broader implications of generating new recipes to allow people selecting food meals based on their favored categories.

Corpus Creation : Annotators & Expert

Asif Ullah Sikder
Annotator -1

Food Industry Expert

Graduate From Malaysia in Hospitality Management

Athban Yusuf
Annotator-2

Owner Kabul Express
TasteBud
Kebab Junction
Food Express
Graduated From Malaysia in Hospitality Management

Nazmus Sakib
Annotator-3

Owner of Tong &
Food Reviewer in Social Media Platform

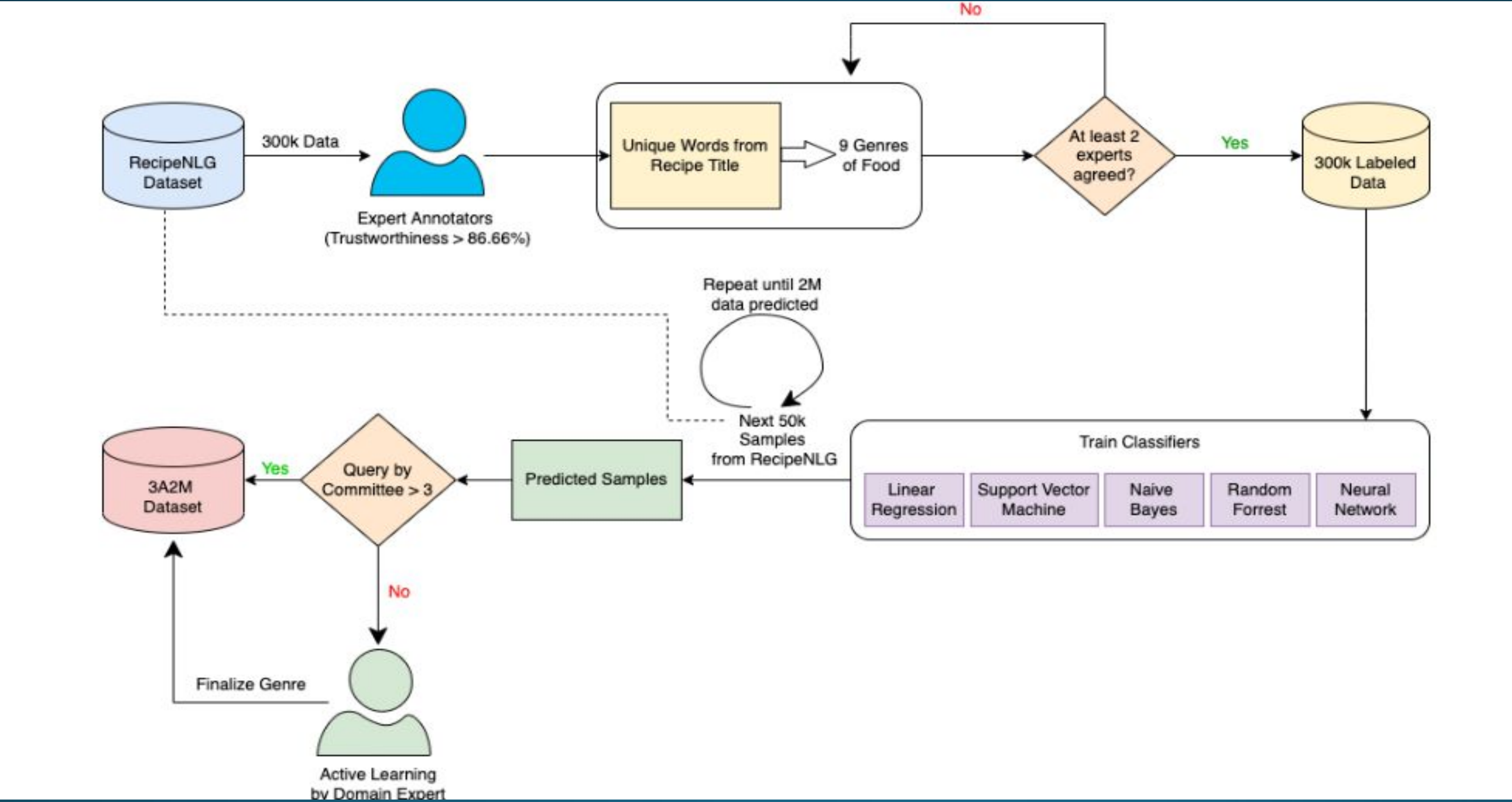
Alamgir Chowdhury
Domain Expert

Executive Chef Six Seasons Dhaka
Faculty Royal Institute of Hospitality Management Dhaka

Corpus Creation : Genre || Classification || Labeling

Genre ID	Genre Name	Facts	No of Instances
1	Bakery	This area mostly contains baked or fried foods that are served in the open or may be stored for a long period.	160712
2	Drinks	Drinks are in the liquid zone and can be blended with any chemical or ionic drink in this zone.	353938
3	NonVeg	This zone includes foods such as curries of poultry, beef, and fish, which can be self-serving or mixed serving.	315828
4	Vegetables	Foods cooked differently than the meats, seafood, and eggs found in this zone. This zone was built just for vegetarians.	398677
5	Fast Food	Only quick food is baked or fried food that cannot be kept for an extended period of time in an open or cold environment.	177109
6	Cereal	Cereals are mainly foods made from corn, wheat, and rice. We have placed meals that are directly generated from grains in the corn zone.	340495
7	Meal	Some items may appear to be quick food, yet they might actually constitute a complete meal.	53257
8	Sides	The medicines, sauces, and toppings are basically sections in the side section.	338497
9	Fusion	Some food that can be properly sorted. Sometimes experts disagree on whether it belongs in a specific category known as fusion meals.	92630

Methodology



Active Learning : Implementation

- To categorize the remaining **17** lac data points, we considered using machine learning to build a model.
- We developed five algorithms for the machine learning process:
  - Logistic Regression
  - Support Vector Machine
  - Naive Bayes
  - Neural Network
  - Random Forest
- The technique follows the Query by Committee procedure, therefore if the result indicated more than three classification algorithms categorizing a work in a given genre, that label was accepted.
- That is, the confidence score is always larger than **60%.**

Outcome : Statistics || Graphs

- Over the first 500 samples, the domain expert assigned 30 control samples. The **trustworthiness scores** of the three annotators are **86.667%, 90.000%, and 96.667%, respectively,** based on their responses. As the set of classes was 9, domain experts felt that more than 80.00 percent would be adequate to qualify the annotators.In this case, the trustworthiness score of each annotator is above 80 percent, and the average trustworthiness score is 90.00 percent, which indicates that the annotators are qualified to annotate.

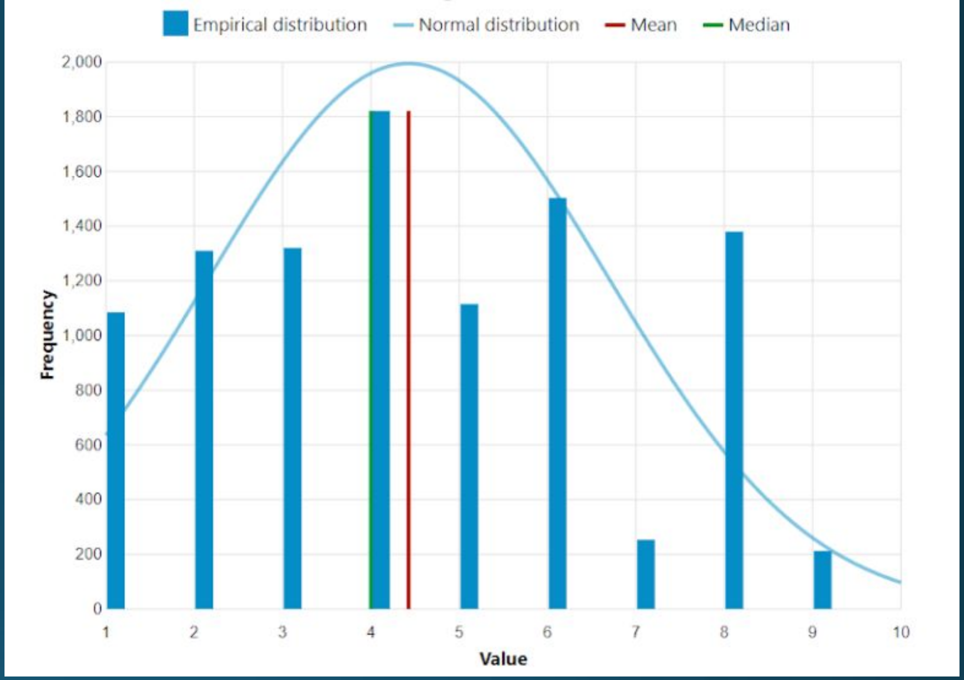
- We have simulated the agreement and found an **IRR value 50.3976667%** for the 3 lac data over 9 genres. As the class levels are high so the IRR value is getting lesser

- After computing over 3 lac annotated data by professionals, we identified the **Fleiss' Kappa value of 0.496527732843039 in the moderate zone.** After processing the data over 3 lac, we identified a Kappa value of 0.5 in the strong zone since there are 9 genres.

- We discovered confidence scores of **100 percent for 89,378 recipes and 66.667 percent for 201,982 recipes across 3 lac entries**.

Outcome : Statistics || Graphs

- We have an Anderson Darlington test Result on **100,000** data distribution.
- If the p-value is less than the significance level, reject the null hypothesis and conclude that the effect is statistically significant.
- P value is **<0.0005.**
- The evidence in our sample is strong enough to reject the null hypothesis at the population level.



Springer EAI RESEARCH MEETS INNOVATION IEEE COMPUTER SOCIETY Bangladesh Chapter http://CenNSER.org CNSER Center for Natural Science & Engineering Research

Conclusion

- We have categorized 300K recipes into nine categories by human annotators and trained five machine learning classifiers to employ active learning for automatically labeling the remaining 1900K instances. In the future, it may be able to normalize them to a specific number of servings.

- Unification of often ambiguous units (e.g. cups, pinch) with respect to the thing they describe, which might have a wide range of applications in and outside of the culinary world, as well as additional unification utilizing knowledge graphs, is another exciting future project.

Future Work

- Because the collection is large and organized by genre, medical sectors, particularly those working with food nutrition, can recommend a variety of meals from it. If the recipe's portion can be estimated, a large area will open up, which is the components calories, which can be used to analyze food calories intake for various types of food analysis or nutrients.

- An application can be created to build a new menu and generate buzz in the food market, giving consumers a new taste and direction to manufacture such delicacies, which may be a big contribution to the culinary sector.

References

1. Bień, M., Gilski, M., Maciejewska, M., Taisner, W., Wisniewski, D., A., L.: Recipenlg: A cooking recipes dataset for semi-structured text generation. In: InProceedings of the 13th International Conference on Natural Language Generation. pp. 22–28 (December 2020)
2. Buneman, P.: Semistructured data. PODS (1997)
3. Cohn, D.A., Ghahramani, Z., MI., J.: Active learning with statistical models. Journal of artificial intelligence research Mar 1(4), 129–45 (March 1996)
4. Cortes, C., Vapnik, V.: Support-vector networks. Machine learning 20(3), 273–297 (1995)
5. Cuisine, S.: Food dictionary. Tech. rep., Online (May 2022), https://www. soscuisine.com/food-dictionary/
6. eathappyproject: Types of cuisines. Tech. rep. (May 2021), https://www.eathappyproject.com/ types-of-cuisines-from-around-the-world-with-their-popular-food/
7. Elo, S., Kääriäinen, M., Kanste, O., Pölkki, T., Utriainen, K., content analysis: A focus on trustworthiness, K.H.Q.: SAGE open 5, 4 (February 2014). https://doi.org/2158244014522633
8. Falotico, R., Quatto, P.: Fleiss' kappa statistic without paradoxes: Quality & quantity 49(2), 463–470 (March 2015)
9. Fisher, M.F.K.: The Anatomy of a Recipe, With Bold Knife and Fork. Counterpoint (1969)
10. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning Data Mining, Inference, and Prediction. Springer Series in Statistics, Springer-Verlag New York, 2nd edn. (2009)
11. Ho, T.K.: Random decision forests. In: Proceedings of the 3rd International Conference on Document Analysis and Recognition. pp. 278–282. Montreal, Q. C. (1995)
12. Kalchbrenner, N., Grefenstette, E., P., B.: A convolutional neural network for modeling sentences". pp. 655–665. Proceedings of ACL (2014)
13. Khodak, M., Saunshi, N., Vodrahalli, K.: A large self-annotated corpus for sarcasm. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation. LREC 2018 (2018)
14. Kiddon, C., Zettlemoyer, L., Choi, Y.: Globally coherent text generation with neural checklist models. In: for Computational Linguistics, A. (ed.) Proceedings of the Conference on Empirical Methods in Natural Language Processing. p. 329–339. Conference on Empirical Methods in Natural Language Processing, Austin, Texas. (2016) 3A2M Cooking Recipes Dataset based on Active Learning 15
15. Kim, G., Kang, S.: Effective transfer learning with label-based discriminative feature learning. Sensors 22(5), 2025 (2022)
16. LeCun, Y., Bengio, Y., G., H.: Deep learning. Nature 521(7553), 436–44 (May 2015)

References

17. Lee, H.H., Shu, K., Achananuparp, P., Prasetyo, P.K., Liu, Y., Lim, E.P., Varshney, L.R.: RecipeGPT: Generative pre training based cooking recipe generation and evaluation system. In Companion Proceedings of the Web Conference 2020 (2020)
18. Mandelbaum, A., Weinshall, D.: Distance-based confidence score for neural network classifiers. Tech. rep. (September 28 2017)
19. Marin, J., Biswas, A., Ofli, F., Hynes, N., Salvador, A., Aytar, Y., Weber, I., A., T.: Recipe1m+: A dataset for learning cross-modal embeddings for cooking recipes and food images. IEEE transactions on pattern analysis and machine intelligenceJul 43(1), 187–203 (July 2019)
20. Mitchell, T.M.: "Machine learning". McGraw-Hill, Maidenhead, U.K., International, ISBN, student edn. (1997), 414 pages
21. Phalippou, L.: The hazards of using irr to measure performance: The case of private equity. SSRN 1111796 (2008)
22. Price, I., Gifford-Moore, J., Fleming, J., Musker, S., Roichman, M., Sylvain, G., Thain, N., Dixon, L., Sorensen, J.: Six attributes of unhealthy conversation. Tech. rep. (October 2020)
23. Razali, N.M., YB., W.: Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. Journal of statistical modeling and analyticsJan 1(2), 1 (2011)
24. Recipes, F.: food.com. Tech. rep., [Online] (February 2022), https://www.food. com/
25. Salvador, A., Hynes, N., Aytar, Y., Mar´ın, J., Ofli, F., Weber, I., Torralba, A.: Learning cross-modal embeddings for cooking recipes and food images. p. 3068–3076. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017)
26. Settles, B.: Active learning literature survey (2009)
27. statology: fleiss-kappa. Tech. rep., Online (March 2022), https://www.statology. org/fleiss-kappa-excel/
28. theodora: Culinary dictionary. Tech. rep., Online (February 2022), https:// theodora.com/food/index.html
29. To, S.H.: Inter-rater reliability irr. Tech. rep. (Mar 2022), https: //www.statisticshowto.com/inter-rater-reliability/
30. wikipedia: List of cuisines. Tech. rep. (July 2021), https://en.wikipedia.org/ wiki/List_of_cuisines
31. Yagcioglu, S., Erdem, A., Erdem, E., Ikizler-Cinbis, N.: Recipeqa: A challenge dataset for multimodal comprehension of cooking recipes. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. p. 1358–1368. EMNLP (2018) (2018)
32. Yang, Z., Blunsom, P., Dyer, C., Ling, W.: Reference-aware language models. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. p. 1850–1859. EMNLP 2017 (2017)

THANK YOU!