Recipe Generation Using Small Language Models

Kanak Yadav (2022611)
BTP Advisor: Dr. Ganesh Bagler

INTRODUCTION

Recipe generation systems offer a smarter alternative to traditional cooking platforms by automatically creating step-by-step instructions from a **given Title and list of Ingredients**. Unlike apps that rely on static databases and complex search tools, this approach enhances user experience through creativity and simplicity.

This project uses **Small language Models** to generate cooking instructions efficiently, avoiding the high computational costs of large models like GPT-2. It enables real-time, domain-specific text generation, ideal for lightweight, user-friendly culinary applications.

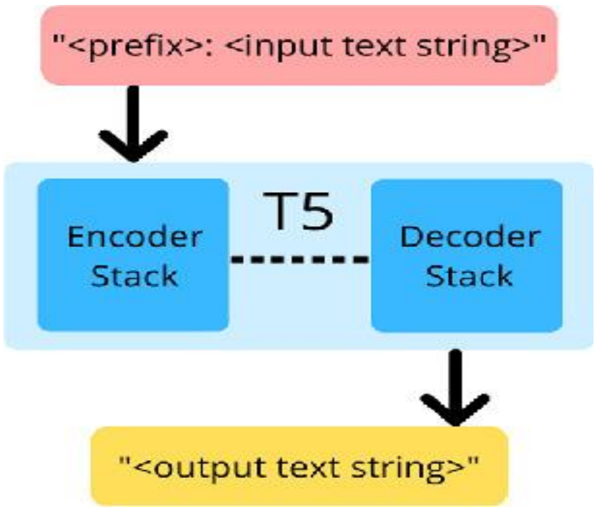
OBJECTIVE

We have two primary objectives.

1. **Instruction Generation:** Given a recipe title and a list of ingredients, the system generates detailed cooking instructions.
2. **Model Comparison:** Evaluates and compares the performance of small language models like T5 Small, DistilBert, and DistilRoBERTA

METHODOLOGY

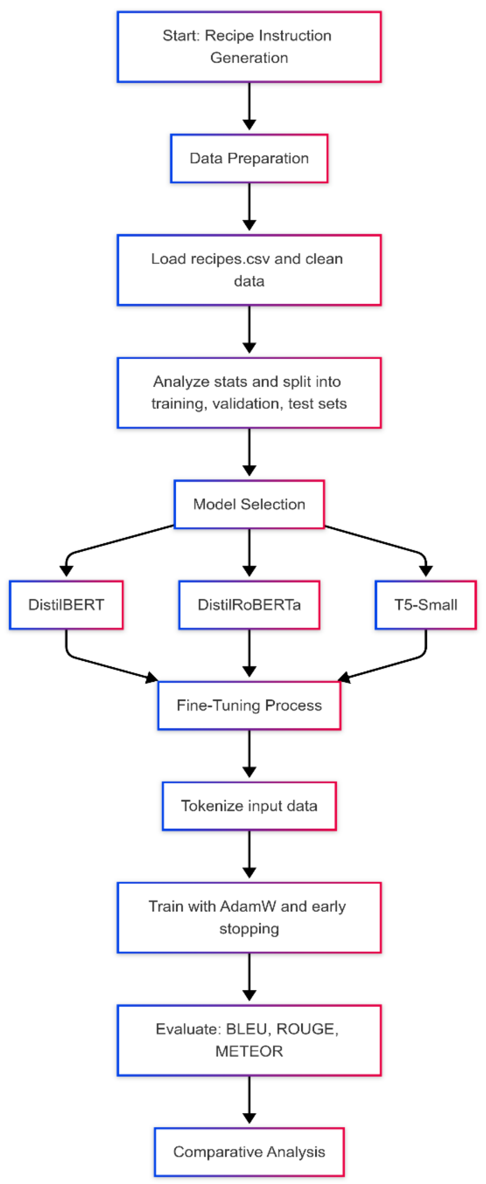
We develop an end-to-end system to generate cooking instructions from recipe titles and ingredients using three compact models—DistilBERT, DistilRoBERTa, and T5-Small. After preprocessing a recipe dataset, we fine-tune each model with task-specific prompts. While encoder-only models (DistilBERT, DistilRoBERTa) are adapted for generation tasks, T5-Small is naturally suited for text-to-text outputs. Model performance is evaluated using both automatic metrics and human review.



METHODOLOGY

Training Method

1. **Input**: Cleaned recipe titles and ingredient lists are provided as input to each language model.
2. **Processing**:
   - Distil-BERT and Distil-RoBERTa encode the input to understand context.
   - T5-Small uses its encoder-decoder architecture to generate instructions in a text-to-text manner.
3. **Output**: Each model generates a set of cooking instructions based on the ingredients and recipe title.
4. **Optimization**: The models are fine-tuned using BLEU, ROUGE, and METEOR-based loss evaluation, optimizing instruction quality, coherence, and fluency

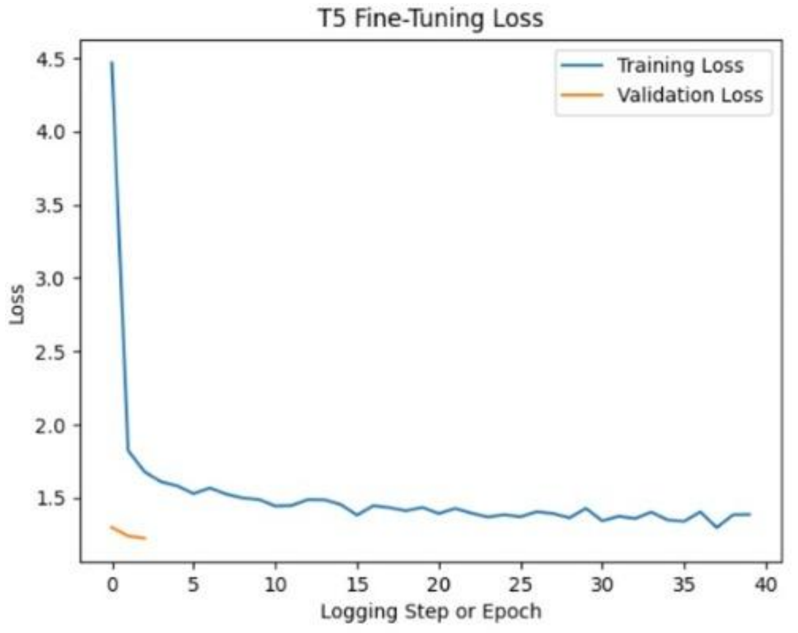


RESULTS

- **T5-Small** achieved the **best overall performance** with BLEU: 6.00%, ROUGE-L: 20.76%, and F1-Score: 24.12%, outperforming encoder-only models due to its **encoder-decoder architecture**, which is well-suited for generative tasks like recipe instruction generation.
- **Distil-RoBERTa** outperformed Distil-BERT, recording BLEU: 7.36%, ROUGE-L: 18.23%, F1-Score: 24.65%, likely due to its **more diverse and extensive pretraining corpus**.
- **Distil-BERT**, while fastest to converge, lagged in generation quality, with BLEU: 2.29%, ROUGE-L: 11.35%, F1-Score: 15.74%.

Model	Architecture	Parameters	Type	Strength	Weakness
DistilBERT	Transformer (Encoder Only)	~66M	Bidirectional Encoder	LightWeight,Fast good at context	Not optimized for text generation
Distil-Roberta	Transformer (Encoder Only)	~82M	Bidirectional Encoder	Improved pretraining from Roberta	Similar to DistilBert
T-5 Small	Transformer (Encoder-Decoder)	~60M	Text-2-Text (seq2seq)	Great for text generations task	Slower inference slightly heavier model

RESULTS



Training and Validation loss

Model	BLEU	ROUGE-L	F1-score
DistilBERT	0.0229	0.1135	0.1574
DistilRoBERTa	0.0736	0.1823	0.2465
T5-Small	0.0600	0.2076	0.2462

Table 5.4: Test-set generation metrics for the three models.

Performance Comparison of Models

CONCLUSION

This study compared three lightweight language models—Distil-BERT, Distil-RoBERTa, and T5-Small—for generating cooking instructions from recipe titles and ingredients. While all models successfully learned domain-specific patterns, their output quality varied. Distil-BERT showed the lowest performance despite fast convergence. Distil-RoBERTa achieved the highest precision and exact matches, making it suitable for tasks requiring accurate phrasing. T5-Small demonstrated the best overall recall and fluency, with the lowest validation loss. Overall, DistilRoBERTa and T5-Small significantly outperformed Distil-BERT, each excelling in different aspects of instruction generation. Distil-RoBERTa excels when exact phrasing matters, while T5-Small is preferable for broader coverage and fluency.

Future: Aims to achieve high-quality recipe Generation with low computational cost, suitable for a limited-resources environment.

REFERENCES

[1] Nair, Rahul Anil, et al. Fine-tuning Language Models for Recipe Generation: A Comparative Analysis and Benchmark Study. arXiv preprint arXiv:2502.02028 (2025). https: //arxiv.org/abs/2502.02028
[2] AI Models FYI. Fine-tuning Language Models for Recipe Generation: A Comparative Analysis and Benchmark Study. https://www.aimodels.fyi/papers/arxiv/fine-tuning-language-mode•
[3] Taneja, Karan, Richard Segal, and Richard Goodwin. Monte Carlo Tree Search for Recipe Generation using GPT-2. arXiv preprint arXiv:2401.05199 (2024). https://arxiv.org/ abs/2401.05199