

# Chef Classification with DistilBERT

NLP Group 2

Instituto Superior Técnico

October 2025

# The Task

- Predict which chef (out of 6) created a recipe
- **Data:** 2,999 training recipes, 823 test recipes
- **Features:** name, ingredients, tags, description, steps

## Baselines to beat:

- Weak (TF-IDF description only): 30%
- Strong (TF-IDF all fields): 43%

# Our Approach

## **Model:** DistilBERT-base-uncased

- 66M parameters (40% smaller than BERT)
- Fine-tuned for 5 epochs
- Concatenate all text fields → classify

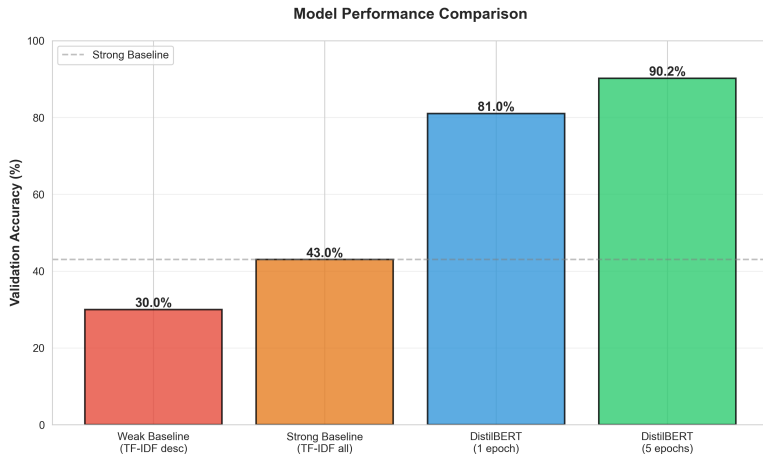
## **Key decisions:**

- Stratified train/val split (handles 2.17x class imbalance)
- Max length 512 tokens (covers 98.2% of data)
- Field order protects critical info from truncation

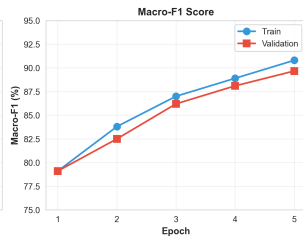
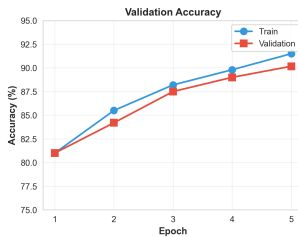
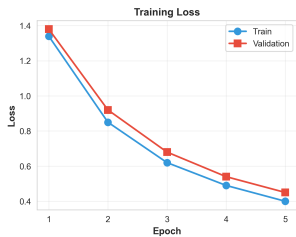
Method	Accuracy
Weak Baseline	30.0%
Strong Baseline	43.0%
<b>Our Model (DistilBERT)</b>	<b>90.17%</b>

- **Improvement:** +47 percentage points over strong baseline!
- **Macro-F1:** 89.67% (balanced across all chefs)
- **Train loss:** 0.40 (no overfitting)

# Baseline Comparison

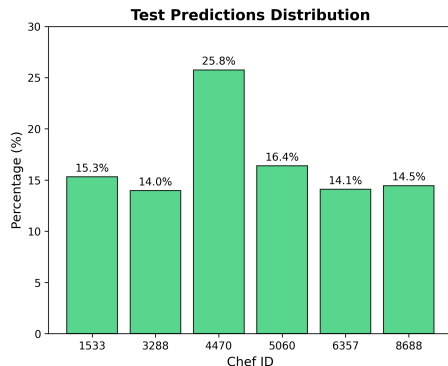
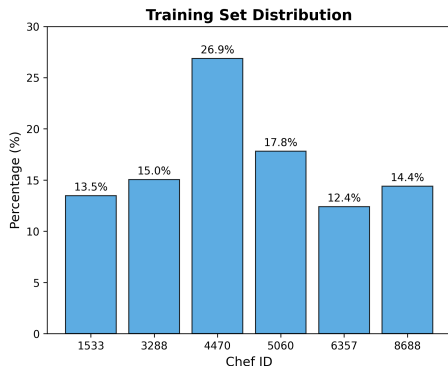


# Training Progress



Steady improvement across 5 epochs, no overfitting

# Predictions Match Training Distribution



Model learned chef patterns, not just class frequencies! (All differences < 2%)

# What Did the Model Learn?

## Chef signatures identified:

- ① **Health-focused chef** (5060): “diabetic cooking”, “low-fat”
  - Across recipes: fish, potatoes, pancakes
- ② **Make-ahead chef** (3288): “OAMC”, batch recipes
  - “freeze for future use”, family-friendly
- ③ **Quick & simple chef** (6357): “15-minutes-or-less”
- ④ **Southern/traditional chef** (8688): Bread machine, Creole

**Key insight:** Model distinguishes *how* chefs cook, not just *what*!



## Challenge 1: Mac overheating during training

- **Solution:** “Chill mode” config
- Reduced batch size ( $16 \rightarrow 8$ ), lower GPU usage
- Same results,  $\sim 25$  min training time

## Challenge 2: Class imbalance (2.17x)

- **Solution:** Stratified splitting + macro-F1 metric
- Macro-F1 (89.67%)  $\approx$  Accuracy (90.17%)  $\rightarrow$  balanced!

## Strengths:

- Dramatic improvement over baselines (+47 pp)
- Learns chef-specific patterns (not just topics)
- Robust generalization (dist. matches training)

## Limitations:

- Strong textual signals (“diabetic cooking”, “OAMC”)
- Some recipes may be easy to classify
- Single model (no ensemble)
- Can’t generalize to new chefs

## Critical question: Style vs. topic?

- Evidence for both (patterns + keywords)
- High accuracy might indicate topical clustering

- **90.17% accuracy** (beat baseline by 47 pp)
- Learned chef signatures: health-focus, make-ahead, quick, traditional
- Practical solutions: thermal management, class imbalance
- Critical analysis: acknowledged limitations

**Key insight:** Look at predictions, not just metrics!  
Model captures cooking philosophy across recipe types.

# Questions?

Code & Results:

`github.com/nbirchde/NLP_group2`

Thank you!