



AraFinNLP 2024: The First Arabic Financial NLP Shared Task

Sanad Malaysha Mo El-Haj Saad Ezzini Mohammed Khalilia Mustafa Jarrar
Sultan Almujaivel Ismail Berrada Houda Bouamor

Task Description

Subtask 1: Multi-dialect Intent Detection



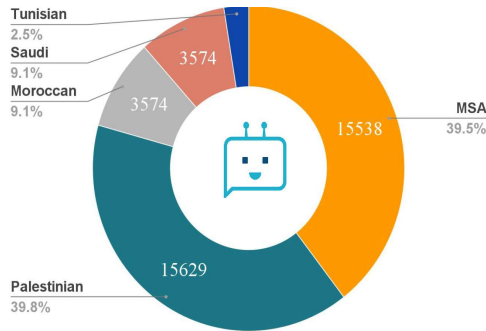
- Participants receive customer queries from a diverse Arabic dialects in the banking domain and classify them into a 77 common intents in the banking domain.

Subtask 2: Cross-dialect Translation and Intent Preservation



- participants translate queries from MSA language to various Arabic dialects, ensuring the preservation of the original intent.

Shared Task Datasets and Evaluation



The ArBanking77 Dataset

| Dialect | Train | Development | Test |
|--------------|-----------------------|--------------|---------------|
| MSA | 10,733 | 1,231 | 3,574 |
| Palestinian | 10,821 | 1,234 | 3,574 |
| Saudi | -- | -- | 3,574 |
| Moroccan | -- | -- | 3,574 |
| Tunisian | -- | -- | 1,000 |
| Sum | 21,554 | 2,465 | 15,296 |
| Total | 39,315 Queries | | |

Evaluation Metrics

Subtask1

- Primary metric: F_1 micro score.
- Secondary metrics: macro precision and recall.

Subtask2

- Primary metric: Bilingual Evaluation Understudy (BLEU).
- Secondary metrics: CHaRacter-level F-score (chrF) and Translation Error Rate (TER).

Shared Task Teams and Results

- 45 unique teams registered for this shared task, with 11 of them actively participated in the test phase.

- 11 teams participated in Subtask 1, while only 1 team participated in Subtask 2.

| Subtask1 Rank | Team | F_1 score |
|---------------|-------------------------------|-------------|
| 1 | MA | 0.8773 |
| 2 | AlexuNLP24 | 0.8762 |
| 3 | BabelBot | 0.8709 |
| 4 | Sultan (no paper submission) | 0.8342 |
| 5 | SemanticCUETSync | 0.8208 |
| 6 | SENIT | 0.8204 |
| 7 | Fired_from_NLP | 0.8014 |
| 8 | Haithem (no paper submission) | 0.7894 |
| 9 | SMASH | 0.7866 |
| 10 | dzFinNlp | 0.6721 |
| 11 | BFCI | 0.4907 |

| Subtask2 Rank | Team | BLEU |
|---------------|------------|-------|
| 1 | AlexuNLP24 | 1.667 |

Open Challenges

- More data needed** to capture a wider range of Arabic dialectal variations.
- Dedicated pre-trained models** for the Arabic banking domain are crucial to improve NLP accuracy.
- More results and **innovative techniques** are needed, necessitating the continuation of the shared task.