UNIVERSITÀ
DI SIENA
1240

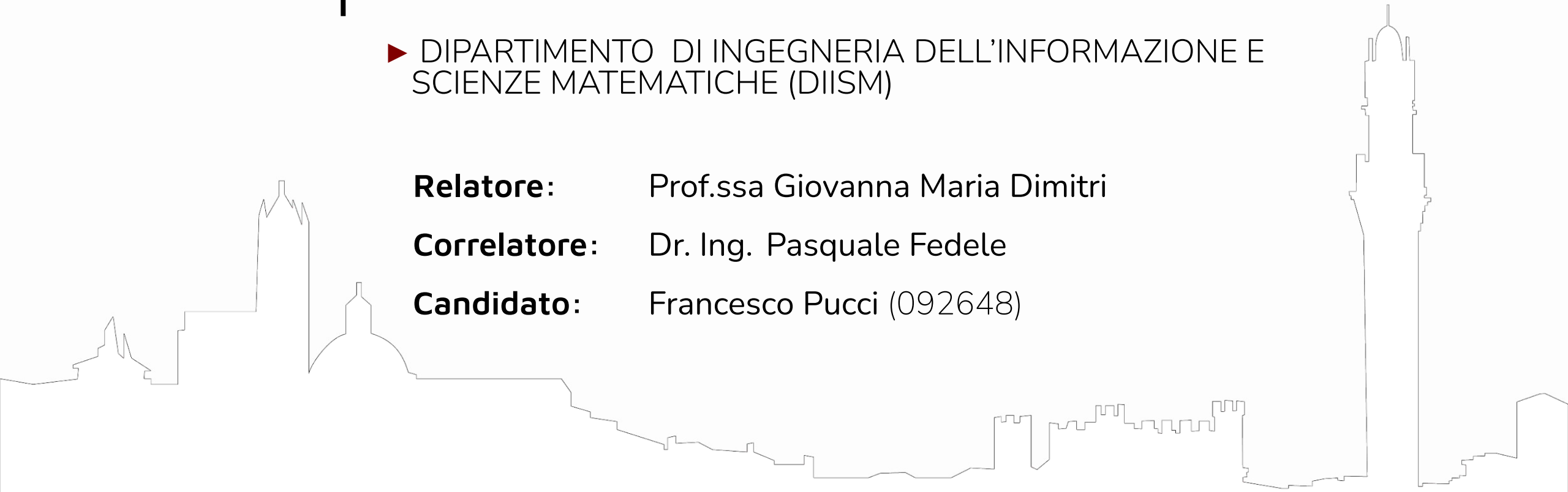
A Speech Emotion Recognition system to perform Sentiment Analysis in a business context

- DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE E SCIENZE MATEMATICHE (DIISM)

Relatore: Prof.ssa Giovanna Maria Dimitri

Correlatore: Dr. Ing. Pasquale Fedele

Candidato: Francesco Pucci (092648)



UNIVERSITÀ DI SIENA
1240

Objective

*The thesis aims to propose a methodology for **sentiment analysis** starting from the **automatic recognition** of emotions in audio samples provided by the company **Blu Pantheon**.*

Blu Pantheon

Business Model

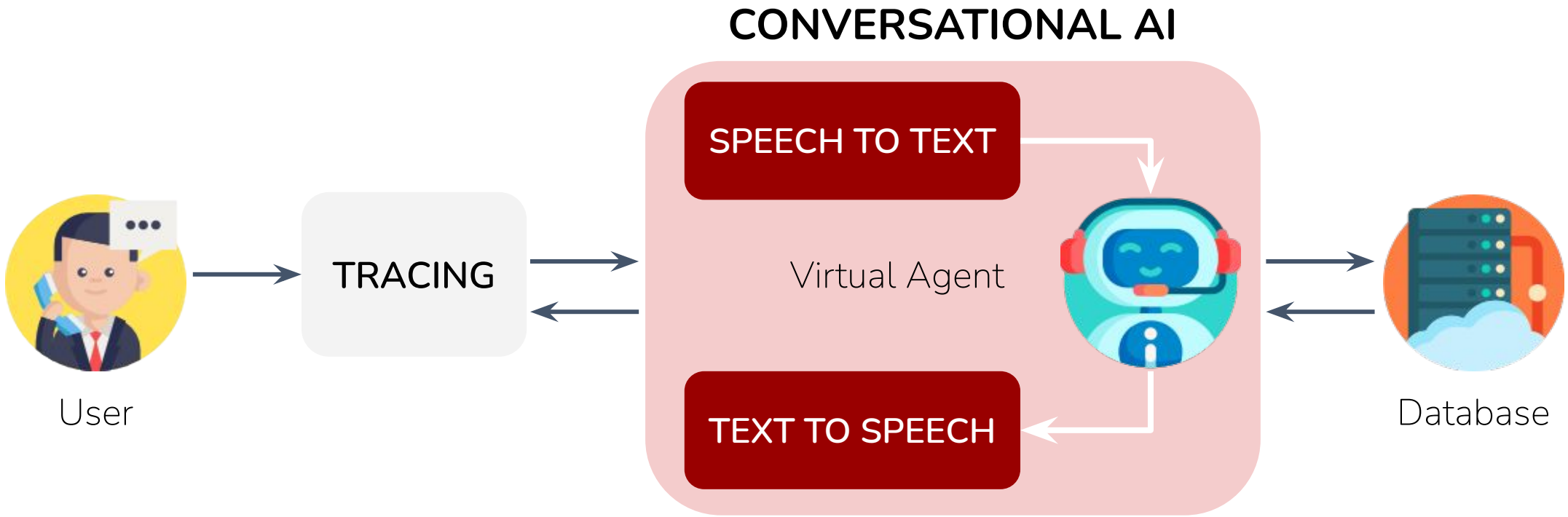
VIRTUAL ASSISTANT

SMART CITY

CYBER SECURITY

Blu Pantheon

Contact Tracing service (COVID-19)



UNIVERSITÀ DI SIENA 1240

Speech Emotion Recognition

"Speech Emotion Recognition (SER) can be defined as extraction of the emotional state of the speaker from his or her speech signal"

Unsupervised context

(3005 audio samples without labels)

UNIVERSITÀ DI SIENA 1240

EMOVO Corpus

Female	Male	Actors	Language
3	3	6	

Sentences	Emotions	Recordings	Duration
14	7	588	0.5 h

UNIVERSITÀ DI SIENA 1240

Data Preparation (EMOVO)

Data Transformation

- **Emotion labels** (Neutral, Disappointment, Fear, Surprise and Sadness)

Data Augmentation

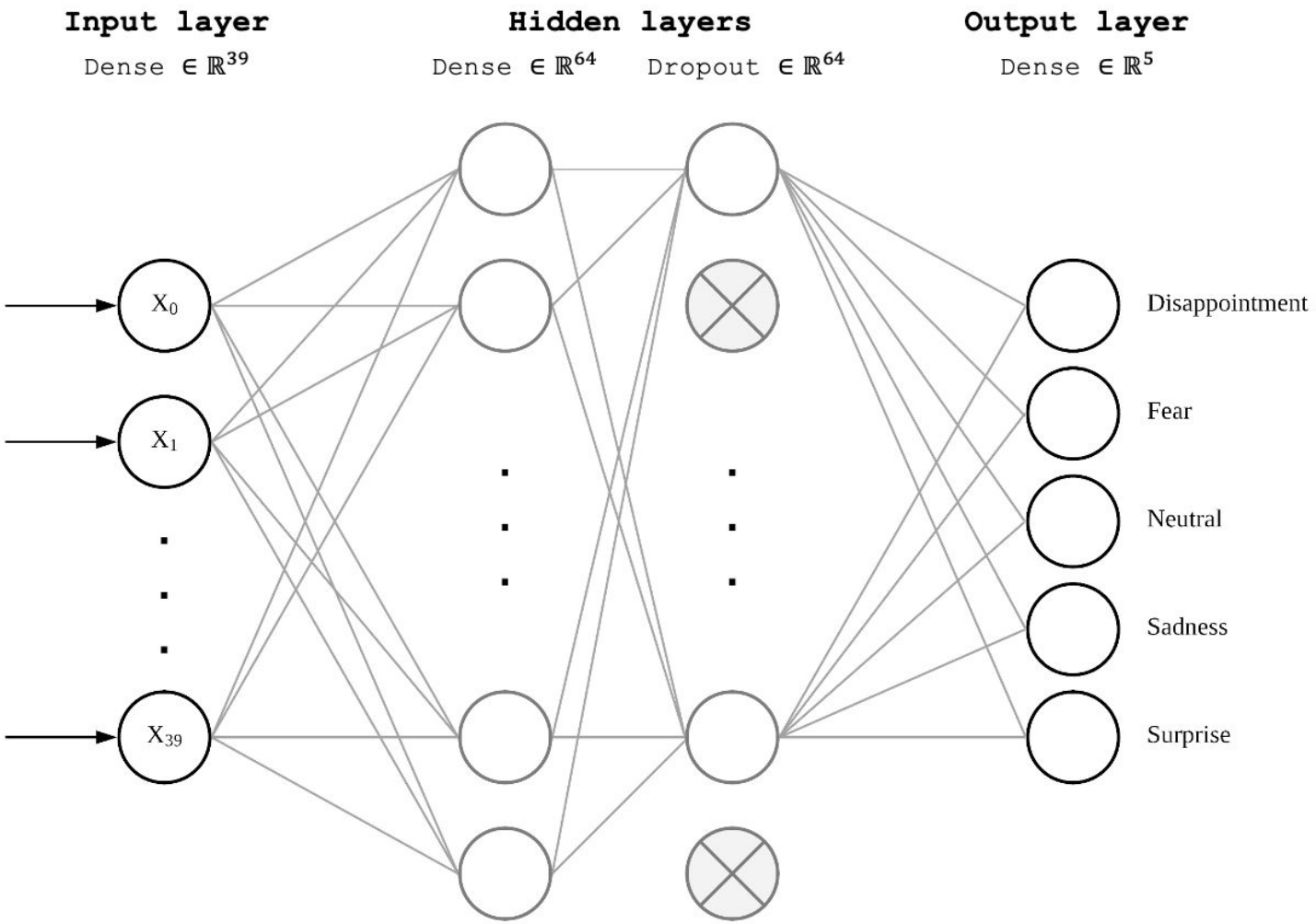
- Noise
- Stretch & Pitch

Feature Extraction

- **Temporal features** (e.g. Energy, ZCR, …)
- **Spectral features** (e.g. Spectral centroid, Spectral roll-off, MFCCs, …)

UNIVERSITÀ DI SIENA
1240

Multilayer Perceptron (MLP)



Model Compile

- Optimizer = `'adam'`
- Loss = `'categorical_crossentropy'`
- Metrics = `'accuracy'`

Model Fit

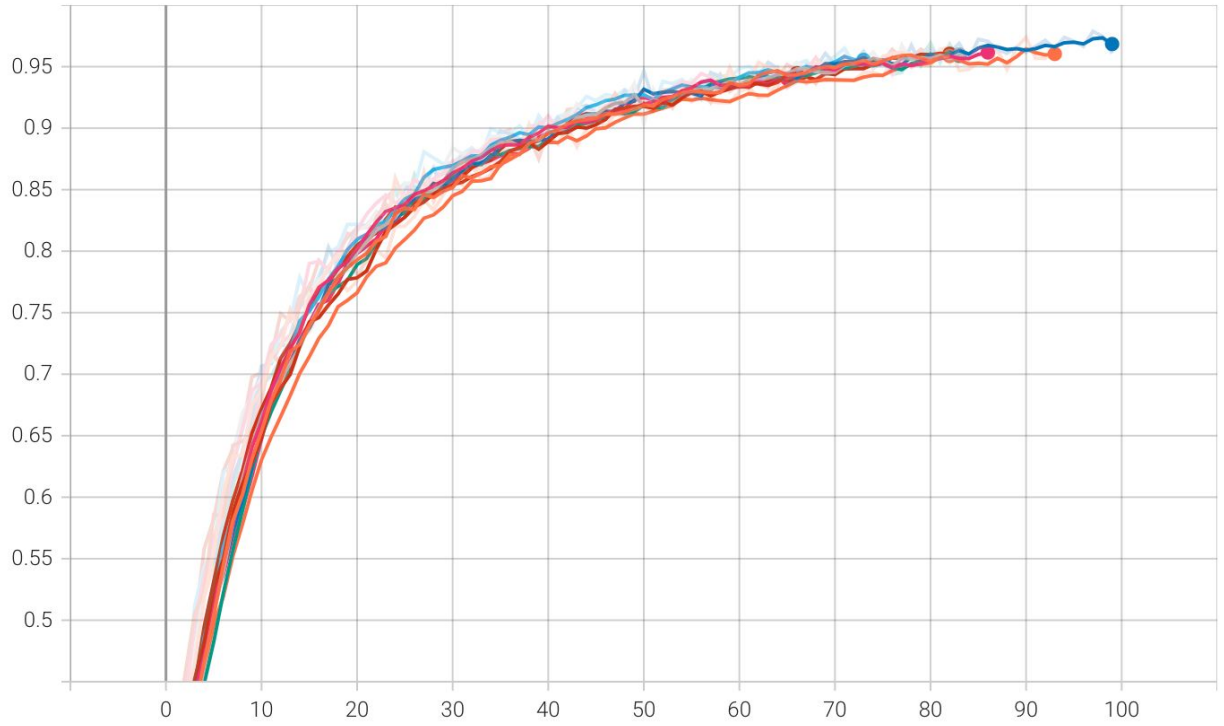
- 10-Fold Cross Validation using 90% of EMOVO dataset
- Epochs = `100`
- Batch size = `64`
- EarlyStopping
  - `patience=10`
  - `min_delta=0.001`
  - `restore_best_weights=True`

UNIVERSITÀ DI SIENA 1240

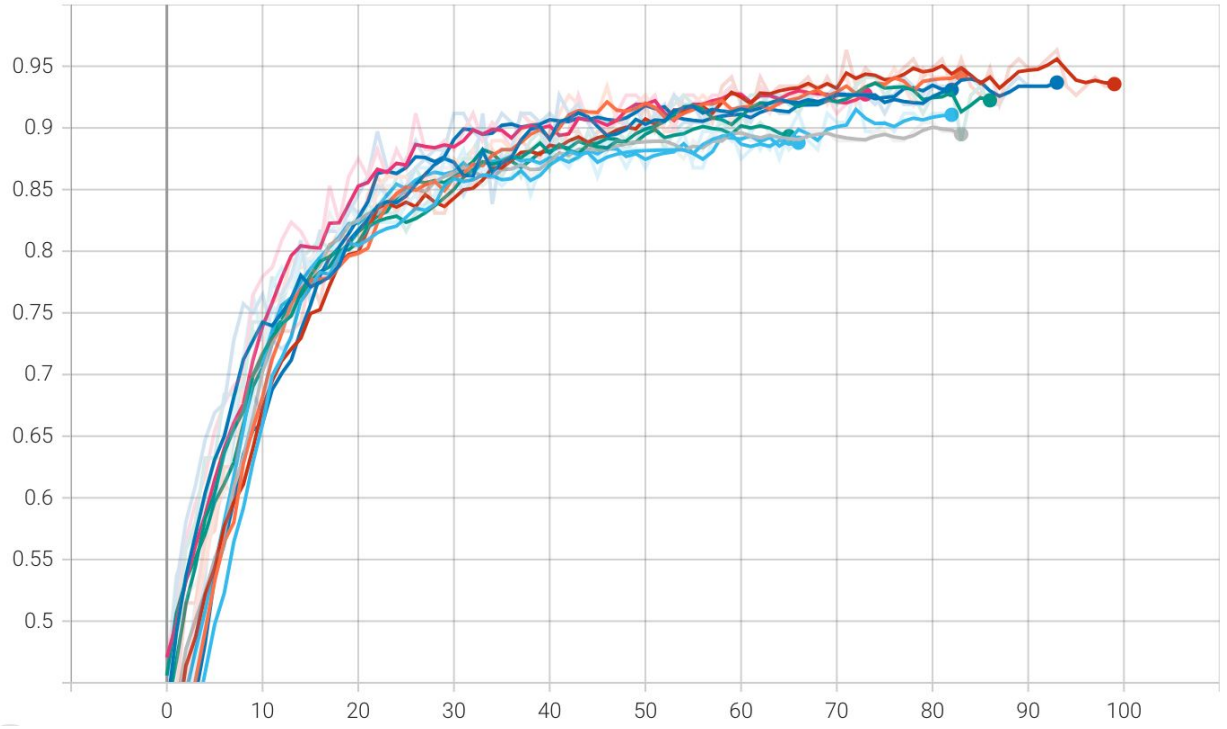
Accuracy (Training vs. Validation)

10-fold Cross Validation → average accuracy of 92.35 %

Training



Validation



UNIVERSITÀ DI SIENA 1240

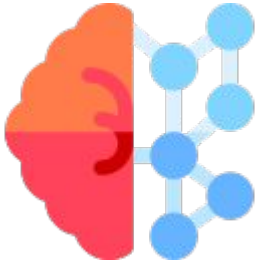
Results (Testing)

	precision	recall	f1-score	support
disappointment	0.89	**0.98**	0.94	52
fear	**0.95**	0.90	0.93	21
neutral	0.88	**1.00**	0.93	28
sadness	**1.00**	0.77	0.87	30
surprise	**0.95**	0.90	0.93	21
macro avg	0.93	0.91	0.92	152
weighted avg	0.93	0.92	0.92	152
accuracy			**0.92**	**152**

UNIVERSITÀ DI SIENA 1240

Transfer Learning

Training the model on **EMOVO Corpus** → Apply it to **Blu Pantheon** experimental dataset



Model performances

Language

Emotions

Durations

Literature

UNIVERSITÀ DI SIENA 1240

Sentiment Analysis

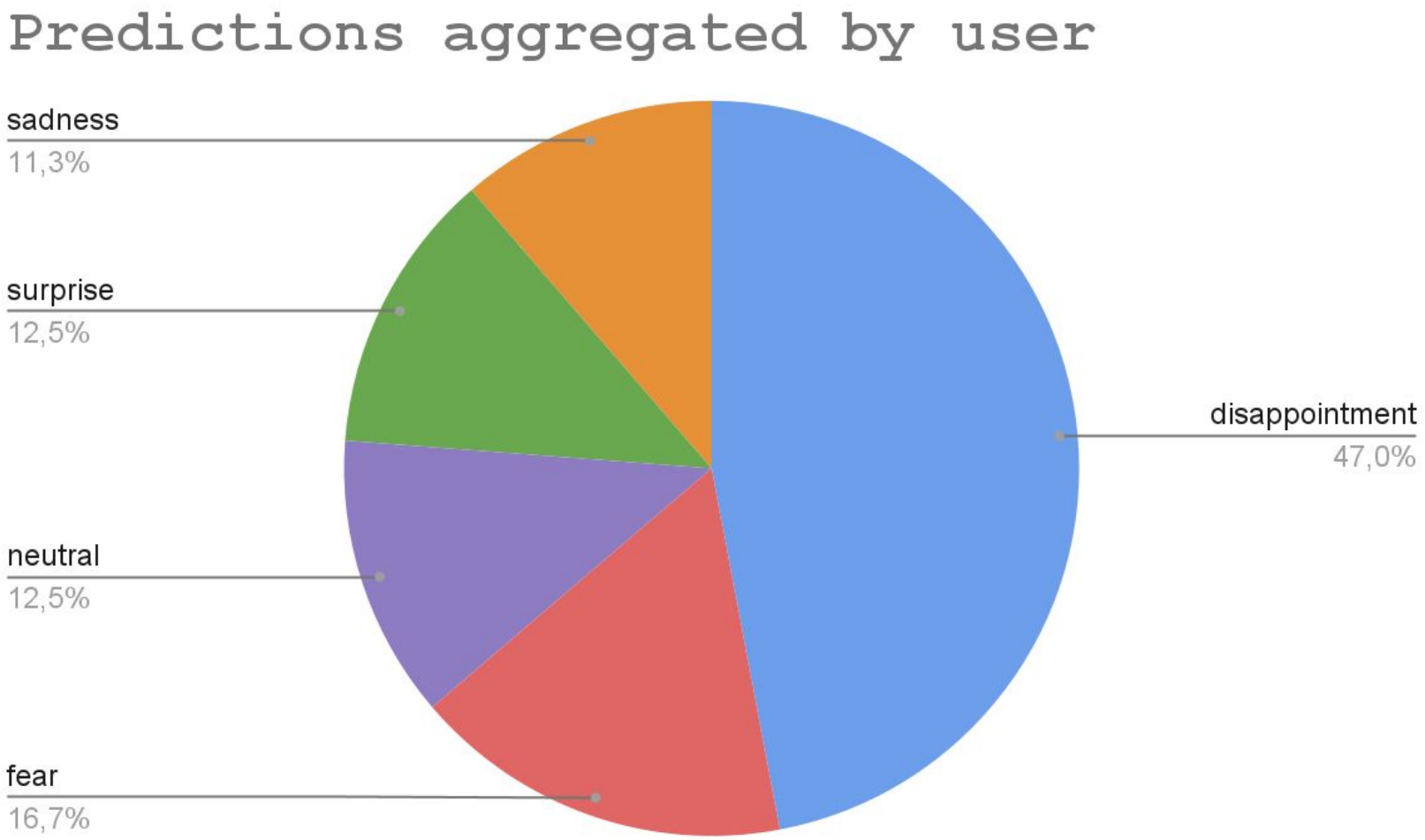
Analysis by user

Analysis by sentence

Empirical analysis

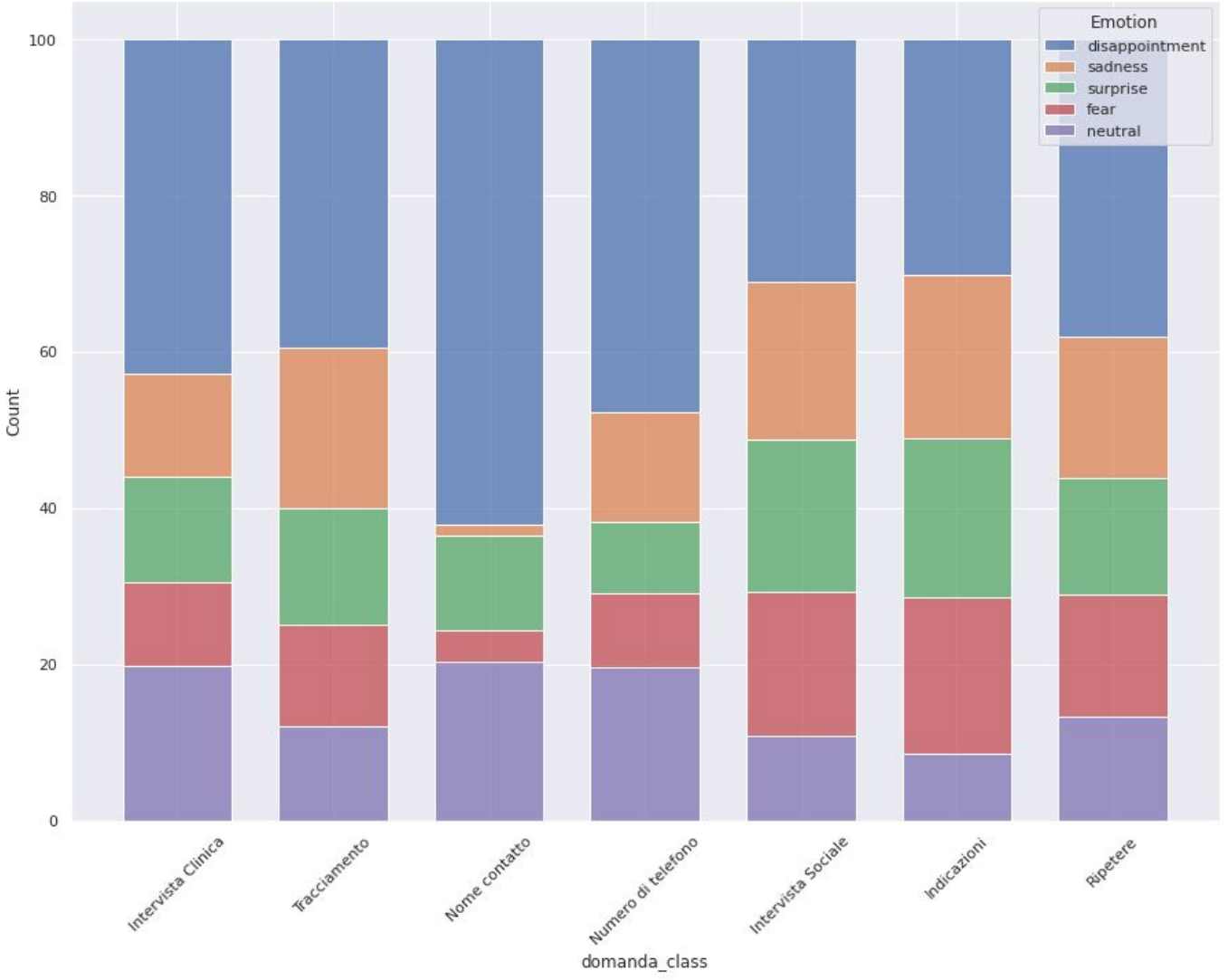
UNIVERSITÀ DI SIENA 1240

Sentiment Analysis (by user)



UNIVERSITÀ DI SIENA 1240

Sentiment Analysis (by sentence)



UNIVERSITÀ DI SIENA 1240

Conclusions and future works

→ Promising results both in training and testing phase (92% accuracy), partially confirmed by the sentiment analysis

→ Create a supervised context (e.g. ask the emotion directly to the user during the conversation or through a final questionnaire)

→ Try using multimodal deep architectures to improve model performance and to allow automatic feature extraction

→ Improve the quality of the service

- Patients with an emotional state at ”risk” can be identified and further assisted
- The conversation flow can be redesigned based on the emotions
- Identification of possible disservices during the calls

UNIVERSITÀ
DI SIENA
1240

Thank you for your attention!