

# Reunião Consórcio

Universidade do Minho  
Hello Project

Braga, 19 de setembro de 2022

# Diagrama de Gantt



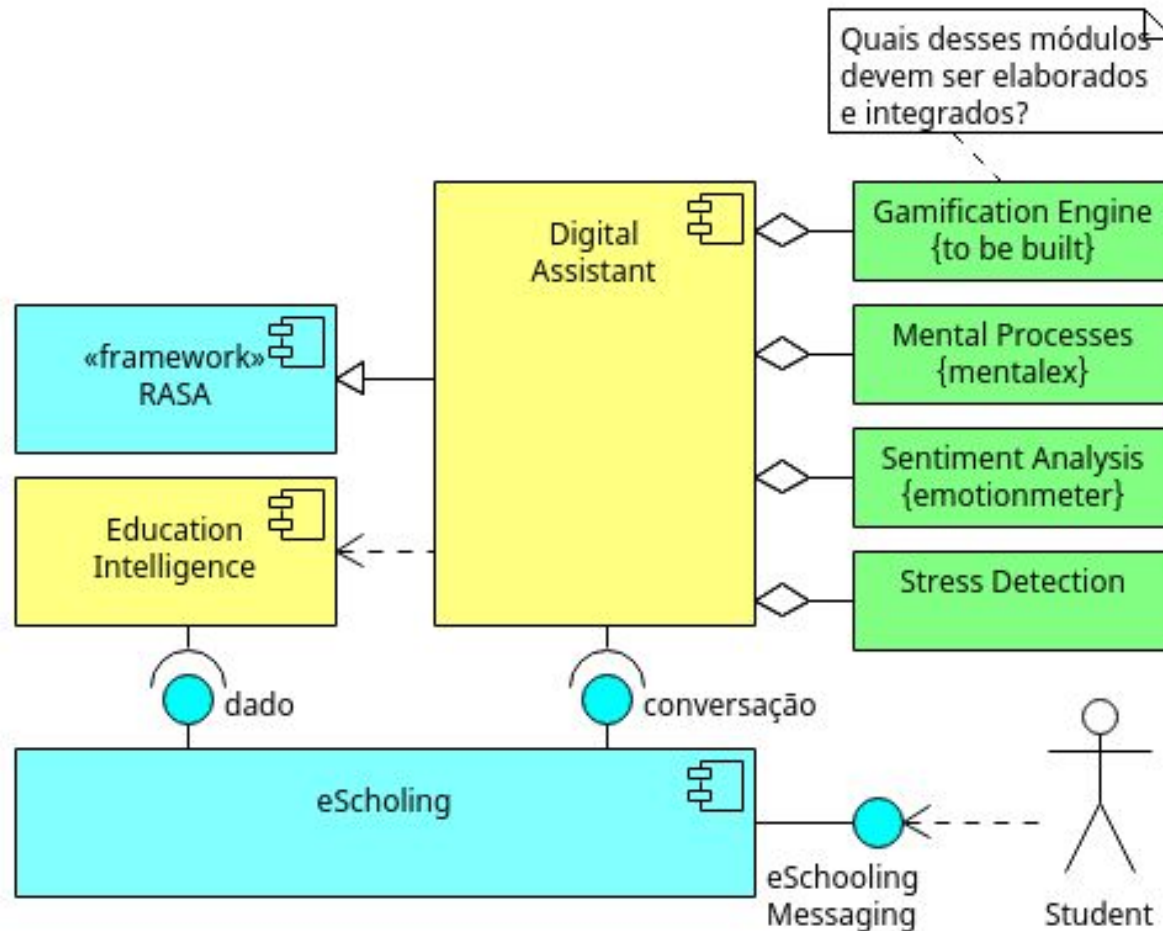
# Atividade 2 – Desenho e Definição da Plataforma

# Tarefa 2.2. Estudo da Integração da Matriz de Risco

- Ainda não recebemos a matriz por parte da UCatólica;
- Estamos a estudar e a integrar protocolos (matrizes) exemplos da literatura.
- Tudo depende deste passo.



# Tarefa 2.3. Definição da Arquitetura e respectivos Módulos



# Atividade 3 – Estudo e Modelação do pré-Processamento dos Dados e Desenvolvimento da Stack Tecnológica para Big Data

# Tarefa 3.1. Profiling

## Preparação das fontes de dados e a extração dos dados Estudo, Compreensão e Visualização dos dados



- Didáxis;
- Escola Secundária Caldas das Taipas;
- Inquéritos realizados nas escolas.

### Inquéritos realizados nas escolas:

- 2º ciclo;
- 200 entradas e 80 colunas;
- 195 entradas podem ser usadas;
- 191 entradas permitem o uso das notas dos alunos;
- Estudar o 3º ciclo.

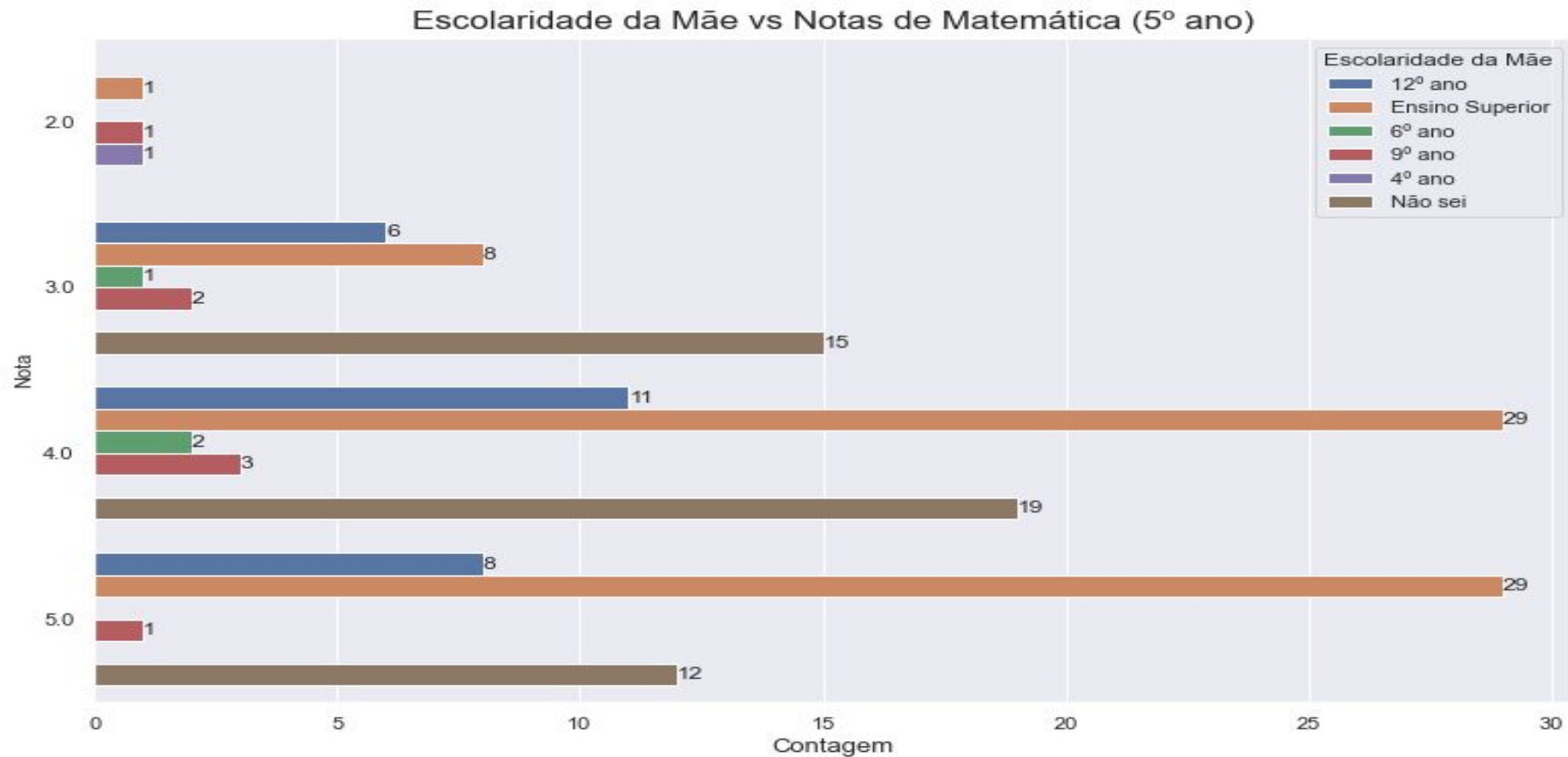
# Tarefa 3.2. Modelação e Preparação e Pré-processamento dos dados

- Manipulação dos valores nulos;
- Standardização;
- Seleção de features (colunas);
- Correção de valores.

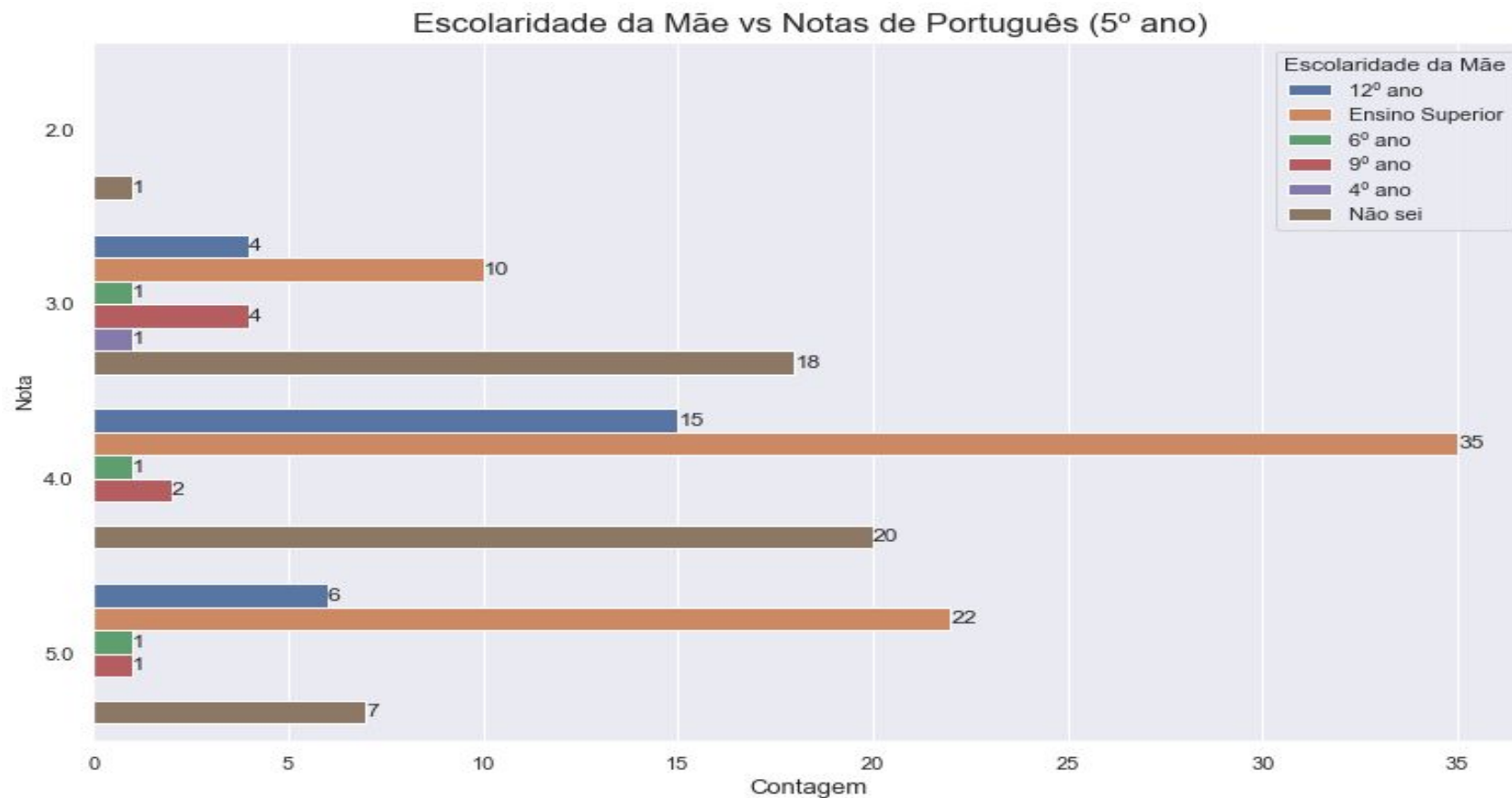




# Tarefa 3.3. Fusão e Integração dos dados e da Informação

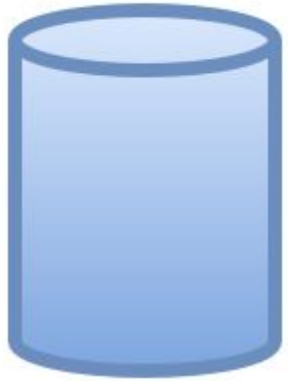


# Tarefa 3.3. Fusão e Integração dos dados e da Informação



# Tarefa 3.4. Aplicação da Tecnologia de Big Data

E-Schooling



# Atividade 4 – Investigação de Modelos de Machine Learning e Deep Learning orientados à Tomada de Decisão

# Tarefa 4.1. Estudo de Técnicas de Computação Afetiva e Análise de Sentimentos

## MentaLex: a Mental Processes Lexicon based on the Essay dataset

Francisco S. Marcondes<sup>1</sup>, Maria Araújo Barbosa<sup>1</sup>, Ricardo Queiroz<sup>2</sup>, Luis Brito<sup>2</sup>, Adelino Gala<sup>3</sup>, and Dalila Durães<sup>1</sup>

<sup>1</sup> Algoritmi Centre, University of Minho, Braga, Portugal

<sup>2</sup> Portucalense University, Porto, Portugal

<sup>3</sup> Digimedia, University of Aveiro, Aveiro, Portugal

francisco.marcondes@algoritmi.uminho.pt, pg42844@alunos.uminho.pt, {44830, 44881}@alunos.upt.pt, adelino@ua.pt, dalila.duraes@algoritmi.uminho.pt

**Abstract.** Considering that the performance of personality predictors are not consistently increasing throughout the years, as an alternative, the idea was to provide big5 personality traits lexicon based on the Essay dataset. The weight of relevance for every trait of each word is calculated by tf/idf on the words in the Essay dataset. However, it was eventually realized that three personalities wordset were overlapped requiring a change of course. The research changed the underlying model from big5 to Vann Joines' mental processes as it appear to suit better with the empirical findings. The resulting lexicon dataset is composed with 3432 commonly used words and 3639 avoided prone words for each mental process. The commonly used words are capable of covering 81% of the Twitter Personality dataset.

**Keywords:** Mental Processes - lexicon - personality - NLP

```
Hey! How was school today?
Your input -> good
Have you done your homework?
Your input -> yess
Great! You are on the path to success. I'm here to
help with whatever you need and to congratulate y
ou on your victories. :)
Your input -> thanks
Your input -> bye
Let me know if you need anything else. Till then,
Goodbye.
```

```
Hey! How was school today?
Your input -> good
Have you study today?
Your input -> no
I'm here to help you. I propose a challenge, You
study for your next assessment and then tell me t
he most interesting subject you learned. What do
you think?
Your input -> ok
I know you can do it.
Let me know if you need anything else. Till then,
Goodbye.
```



# Tarefa 4.2. Investigação dos Modelos de Análise Preditiva

## EduBot: A Proof-of-Concept for a High School Motivational Agent\*

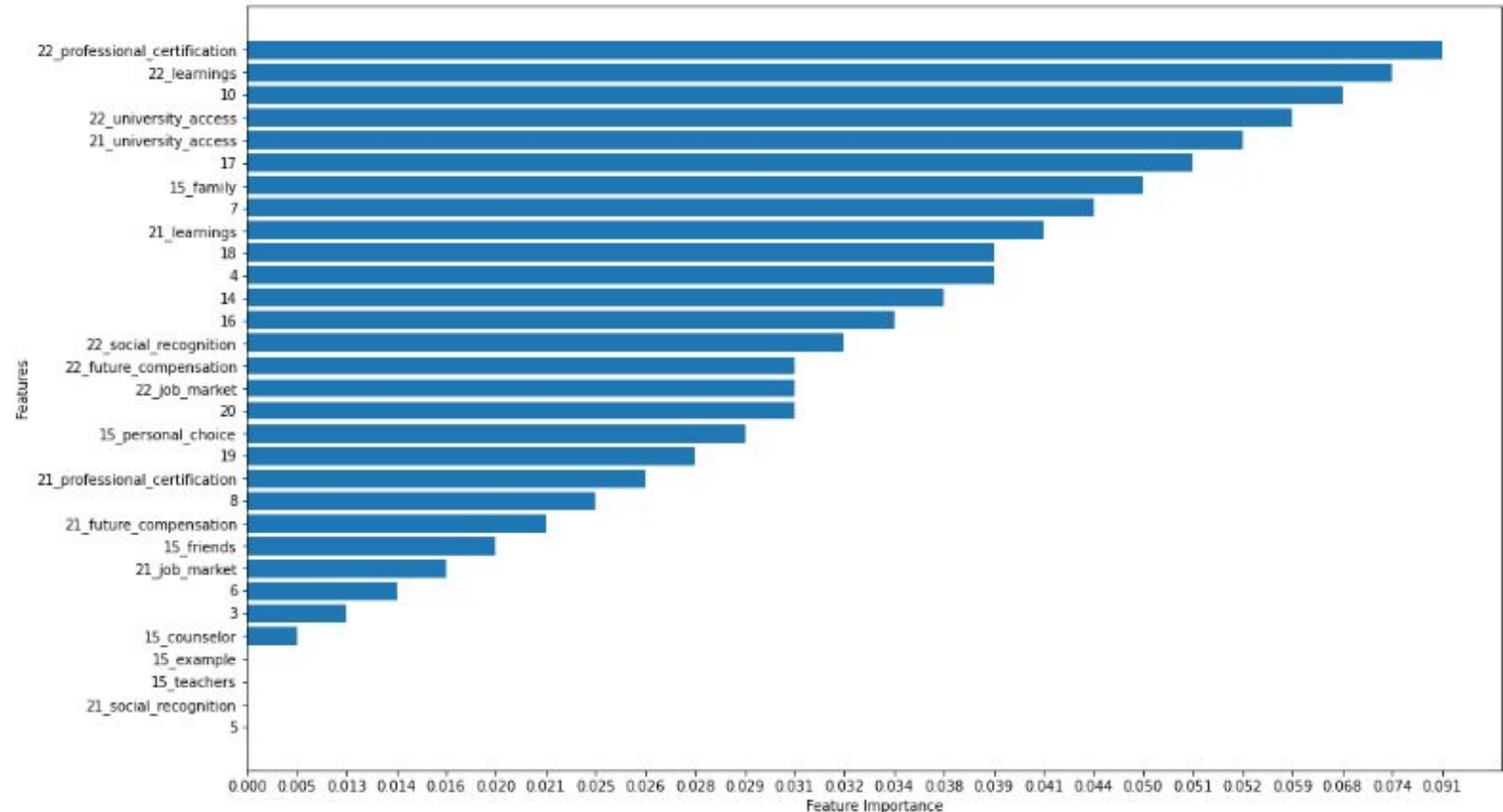
Hugo Faria<sup>1</sup>[0000-0003-2223-7685], Maria Araújo Barbosa<sup>1</sup>[0000-0002-0464-0991],  
Bruno Veloso<sup>1</sup>[0000-0003-3243-1633], Francisco S.  
Marcondes<sup>1</sup>[0000-0002-2221-2261], Celso Lima<sup>2</sup>[0000-0002-4338-2059], Dalila  
Durães<sup>1</sup>[0000-0002-8313-7023], and Paulo Novais<sup>1</sup>[0000-0002-3549-0754]

<sup>1</sup> ALGORITMI Centre, University of Minho, Braga, Portugal  
(a81283, pg42844, a78352)@alunos.uminho.pt, (francisco.marcondes,  
dalila.duraes)@algoritmi.uminho.pt, pjon@di.uminho.pt

<sup>2</sup> High School of Caldas das Taipas  
diretor@esct.pt

**Abstract.** Motivation appears to play a role in student dropout. This paper uses a dataset with high school information (756 entries) for creating a model that relates year failures with the other features on the dataset. The resulting model, based on XGBoost, shows that the top-ranked features are related with motivation issues corroborating with the initial observation. In this sense, a digital assistant embedded with a motivation module may aid on improving motivation and on avoiding dropout. In other words, if the predictor detects year failure possibility it start to act on motivating the student. Considering this paper being direct to underage students, this paper stops on the proof-of-concept level considering the actual dialogues and live tests are expected to be performed with the support of educational psychologists.

**Keywords:** Educational Data Mining · Motivational Assistant · Predict At-Risk Students · Student's Performance



# Tarefa 4.3. Desenvolvimento do Módulo de Suporte à Tomada de Decisão

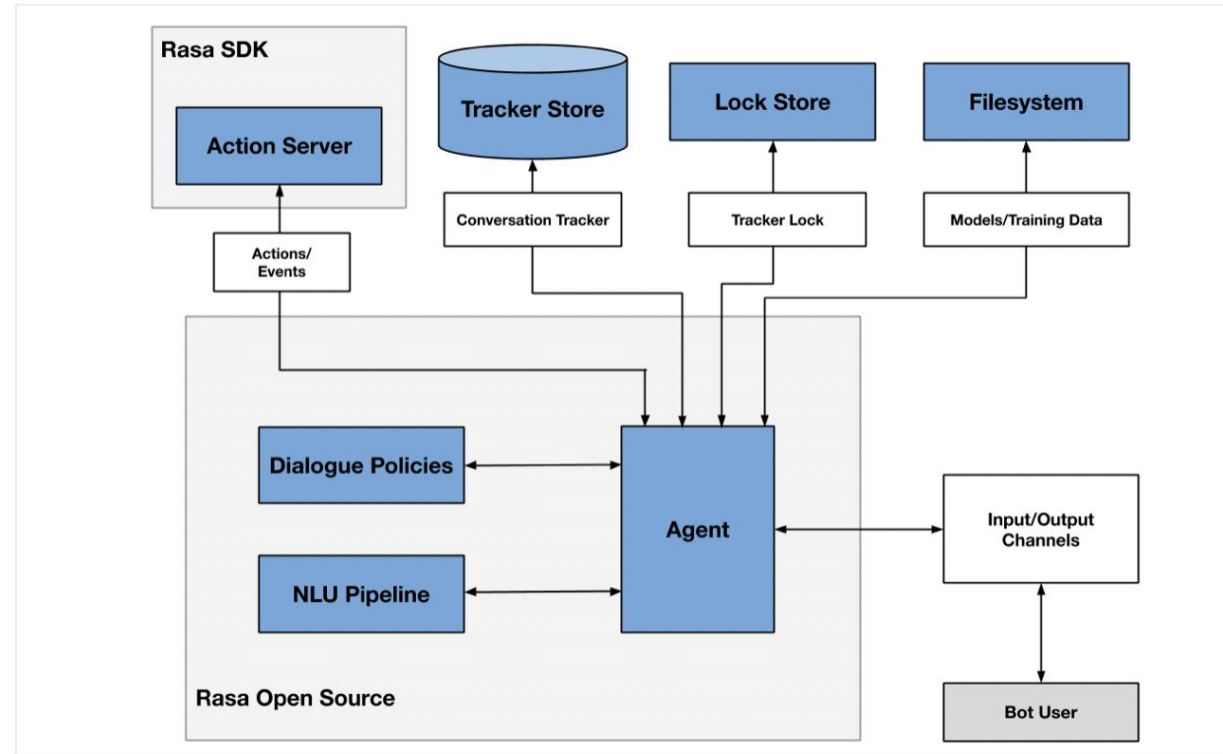
**a depender da matriz de riscos**



# Atividade 5 – Desenvolvimento do Tutor Virtual e de Assistência ao Aluno



# Tarefa 5.1. Desenho da Arquitetura



[Fonte](#)

# Tarefa 5.2. Desenvolvimento da camada de Machine Learning e de Deep learning

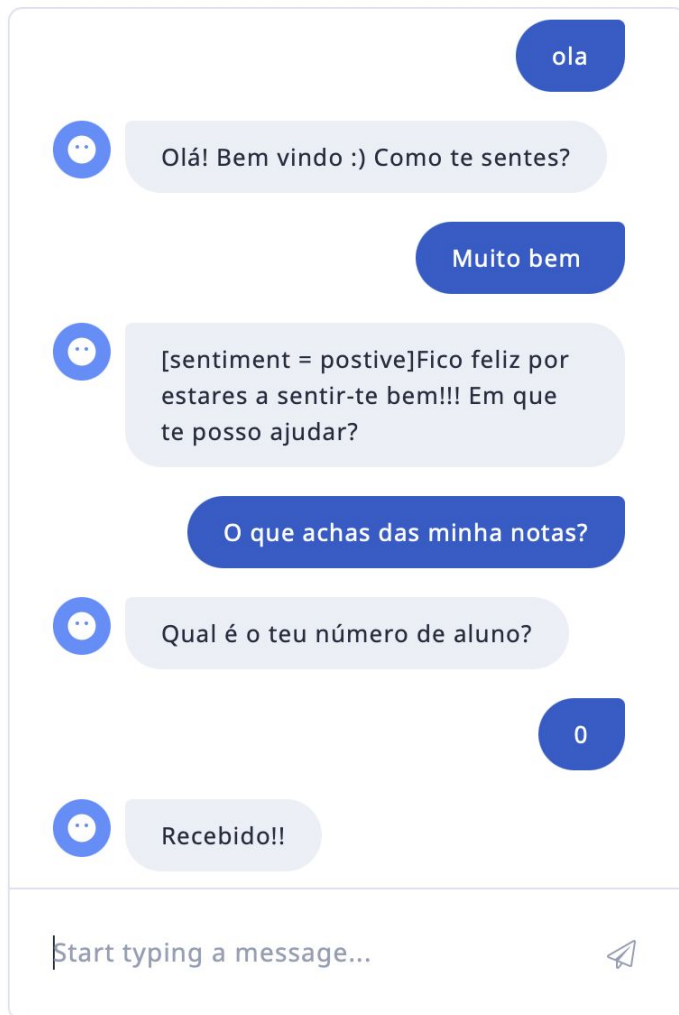
```
language: pt

pipeline:
- name: WhitespaceTokenizer
- name: RegexFeaturizer
- name: LexicalSyntacticFeaturizer
- name: CountVectorsFeaturizer
- name: CountVectorsFeaturizer
  analyzer: char_wb
  min_ngram: 1
  max_ngram: 4
- name: DIETClassifier
epochs: 100

policies:
- name: MemoizationPolicy
- name: TEDPolicy
  max_history: 5
  epochs: 100
- name: RulePolicy
```

**Grau de Maturidade:** Prova de Conceito

# Tarefa 5.3. Desenvolvimento da Camada de Dados e da Camada de Apresentação



ola

Olá! Bem vindo :) Como te sentes?

Muito bem

[sentiment = positive]Fico feliz por estares a sentir-te bem!!! Em que te posso ajudar?

O que achas das minha notas?

Qual é o teu número de aluno?

0

Recebido!!

Start typing a message...

**Camada de Apresentação:** Prova de Conceito  
**Camada de Dados:** N/A



Relatório



**Universidade do  
Minho**  
Escola de Engenharia  
Departamento de  
Informática

# Reunião Consórcio

Universidade do Minho  
Hello Project

Braga, 19 de setembro de 2022

