

Projeto ED - Coleta de dados.

- 1) Coletar noticias utilizando uma api livre para isso (<https://gnews.io/>)
- 2) Para cada título, utilizar uma API de processamento de sentimento (<https://www.meaningcloud.com/>), e coletar o sentimento do título da noticia
- 3) De preferência as duas apis não devem ser chamadas no mesmo coletor (Baixe as notícias primeiro, salve-as, e depois colete os sentimentos)
- 4) Salvar a informação (título | sentimento) em parquet, salvar mais informação é um diferencial (em especial: fonte e pais)
- 5) Escrever uma dag airflow (ou um código do cron), para as coletas necessárias.
- 6) Seu coletor deve ter tolerância a erros do servidor (5xx)
- 7) As api's sugeridas não são obrigatórias (você pode optar por outra api de notícias, e outra api de análise de sentimento). Porém elas obrigatoriamente devem ter um "free tier"
- 8) Cuidado com os limites do free-tier das api's (ambas tem limites, mas não precisam de informação de pagamento para o cadastro). Idealmente realize testes com dados "mock"
- 9) Em caso de problemas com as Apis, você pode desenvolver api's próprias.
- 10) Coletar de pelo menos 2+ países/línguas distintos é um diferencial.
- 11) Grupos de 1 a 4 pessoas.
- 12) Todos os membros do grupo devem entregar individualmente (mas o nome no trabalho deve ser de todo)
- 13) Entregas via git ou email: alvaroazucchi@gmail.com
- 14) Entregue somente na sexta (10/03/2023). Haverá um forms para oficialização da entrega.