# ****Proposta de Projecto Final de Machine Learning****

## 1. Título do Projecto

**Plataforma Digital Colaborativa para Promoção do Turismo e Património Cultural**

## Resumo da Ideia

Este projecto propõe o desenvolvimento de uma **plataforma digital colaborativa** que visa promover o turismo e o património cultural através de uma **rede social inteligente**, onde os utilizadores podem:

* Publicar experiências de viagem;
* Partilhar recomendações e informações sobre locais turísticos;
* Avaliar destinos e serviços;
* Aceder a bibliotecas sobre biodiversidade e pontos históricos;
* Conectar-se a agências de viagem e prestadores de serviços.

A componente de **Machine Learning** será integrada para:

1. **Classificar publicações** (tipo de post: recomendação, feedback, informação cultural, etc.);
2. **Recomendar conteúdo relevante** (baseado nos interesses e histórico dos utilizadores);
3. **Detetar conteúdo irrelevante ou spam**, melhorando a qualidade da informação partilhada.

O sistema contribuirá para a digitalização do setor do turismo e para a valorização do património cultural nacional, promovendo uma comunidade virtual colaborativa e sustentável.

## 3. Revisão de Literatura

A aplicação de **Inteligência Artificial (IA)** e **Machine Learning (ML)** no setor do turismo tem crescido exponencialmente nos últimos anos.  
Estudos recentes destacam três áreas principais de impacto:

### 3.1 Classificação automática de conteúdo

Segundo Zhang et al. (2022), a utilização de modelos de **Processamento de Linguagem Natural (NLP)**, como TF-IDF, Word2Vec e BERT, permite classificar e analisar automaticamente grandes volumes de texto em plataformas online, como avaliações de viagens, blogs e redes sociais.

### 3.2 Sistemas de recomendação

De acordo com Ricci et al. (2021), os **sistemas de recomendação baseados em filtragem colaborativa e filtragem de conteúdo** têm sido amplamente aplicados em plataformas como TripAdvisor e Booking.com para sugerir destinos, hotéis ou experiências personalizadas aos utilizadores.

### 3.3 Deteção de spam e conteúdo irrelevante

Gupta & Kumar (2023) demonstram que a utilização de classificadores supervisionados (como Naive Bayes e Random Forest) combinados com análise semântica é eficaz para identificar conteúdo duplicado, publicações automatizadas e linguagem ofensiva em redes sociais.

Esses estudos mostram que o uso de ML e NLP em plataformas digitais não apenas melhora a experiência do utilizador, mas também contribui para a integridade das comunidades online e a eficiência na partilha de conhecimento.

## 4. Dados

Como o projeto é inovador e baseado num contexto angolano, os dados iniciais serão **simulados** e posteriormente **coletados de forma colaborativa** através da própria plataforma.

### 4.1 Fontes de dados

* Publicações e avaliações dos utilizadores (texto, imagens, data, localidade);
* Metadados (número de likes, comentários, partilhas);
* Feedback sobre destinos e serviços turísticos;
* Dados externos opcionais (descrições de locais históricos, pontos turísticos, e biodiversidade — obtidos de APIs públicas como Google Places ou Wikipedia).

### 4.2 Estrutura dos dados

| **Campo** | **Tipo** | **Descrição** |
| --- | --- | --- |
| id\_post | Inteiro | Identificador único da publicação |
| texto | Texto | Conteúdo escrito pelo utilizador |
| categoria | Texto | Tipo de publicação (recomendação, feedback, história, etc.) |
| likes | Inteiro | Número de curtidas |
| comentários | Inteiro | Número de comentários |
| spam | Booleano | Indica se o conteúdo é considerado spam |
| user\_id | Inteiro | Identificador do utilizador |
| data\_postagem | Data | Data e hora da publicação |

## 5. Tecnologias a Utilizar

### Linguagens e Ferramentas

| **Componente** | **Tecnologia** |
| --- | --- |
| **Frontend** | HTML, CSS, JavaScript |
| **Backend/API** | Django(Python) |
| **Banco de Dados** | PostgreSQL |
| **Machine Learning** | Scikit-learn, TensorFlow, NLTK, SpaCy |
| **Visualização e Dashboard** | Plotly, Matplotlib |
| **Gestão de dependências** | Pip / Virtualenv |

### Modelos de ML

| **Objetivo** | **Técnica** | **Modelo** |
| --- | --- | --- |
| Classificação de posts | NLP + TF-IDF / Word2Vec | Logistic Regression, Naive Bayes |
| Recomendação de conteúdo | Filtragem colaborativa / Conteúdo | KNN, Recommender híbrido |
| Deteção de spam | NLP + Regras + Classificação binária | Random Forest, SVM |

## 6. Resultados Esperados

* Um sistema capaz de classificar automaticamente o tipo de post;
* Um modelo de recomendação de conteúdo relevante;
* Um detector de spam eficaz;
* Uma interface interativa para utilizadores e administradores;
* Contribuição tecnológica para o setor do turismo e cultura;
* Relatório técnico e vídeo de demonstração do sistema em funcionamento.

## Cronograma Resumido

| **Etapa** | **Descrição** | **Data Limite** |
| --- | --- | --- |
| 1 | Proposta de ideia, revisão de literatura, dados e tecnologia. | 20 de Outubro |
| 2 | Preparação de dados / Engenharia de recursos / Plano de implantação. | 24 de Outubro |
| 3 | Documentação do projeto de ML e da implementação. | 31 de Outubro |
| 4 | Slides e documentação final. | 5 de Novembro |

**OBS.:** As linguagens escolhidas, bem como os modelos, podem sofrer alteração na implementação do projecto.