

PROJECTO FINAL DO BOOTCAMP FTL UNDP - ANGOLA 2025

GRUPO 1

MEMBROS DO GRUPO:

1. Arsénio Eurico Muassangue
2. Edgar Delfino Tchissingui
3. Francisco Adão Vika Manuel
4. Raquel de Jesus

SECÇÃO I - PROPOSTA DE IDEIA

TÍTULO DO PROJECTO: **Recomendador Inteligente de Hospedagem Sustentável** *(Alinhado aos ODS 8, 12 e 13)*

A. CONTEXTUALIZAÇÃO DO PROBLEMA

O sector do turismo em Angola representa uma das maiores oportunidades de diversificação económica e desenvolvimento sustentável do país, particularmente nas províncias com elevado potencial natural e cultural, como Namibe, Benguela, Huíla, Bié e Cuando Cubango. Contudo, o modelo actual de hospitalidade ainda enfrenta um problema estrutural: a invisibilidade dos alojamentos sustentáveis, pequenas pousadas, lodges e quintas ecológicas que praticam gestão ambiental responsável, mas permanecem fora dos grandes circuitos digitais.

Em plataformas internacionais como Booking ou TripAdvisor, a maioria dos estabelecimentos listados é avaliada apenas com base em preço, conforto ou localização, sem considerar indicadores de sustentabilidade (uso de energia limpa, gestão de resíduos, impacto social local). Isso gera uma lacuna crítica: os viajantes conscientes não conseguem identificar facilmente opções verdes, enquanto empreendedores comprometidos com práticas ecológicas têm pouca visibilidade e reconhecimento.

Segundo a **UNWTO (2023)**, o turismo sustentável deverá representar mais de 60% das preferências globais de viagem até 2030, impulsionado por políticas climáticas e mudanças comportamentais dos consumidores. A

UNDP (2022) destaca ainda que a integração de tecnologias digitais e inteligência artificial é essencial para transformar o turismo africano em um vector de crescimento verde e inclusivo. Em Angola, onde o turismo ecológico tem potencial inexplorado, falta um ecossistema digital que conecte sustentabilidade, inovação e desenvolvimento local.

A ausência de sistemas inteligentes de recomendação faz com que alojamentos ambientalmente responsáveis permaneçam invisíveis ao turista médio, limitando sua competitividade e dificultando o alinhamento do país às metas de sustentabilidade global. Essa lacuna digital é, simultaneamente, um desafio e uma oportunidade: a oportunidade de posicionar Angola como pioneira em turismo inteligente e sustentável na África Austral.

B. JUSTIFICATIVA

A proposta do *Recomendador Inteligente de Hospedagem Sustentável* responde directamente às metas dos Objectivos de Desenvolvimento Sustentável (ODS):

- **ODS 8 — Trabalho Decente e Crescimento Económico:** ao promover a visibilidade de alojamentos sustentáveis e de base comunitária, o projecto impulsiona o turismo verde, cria empregos qualificados e fomenta microempreendimentos locais.
- **ODS 12 — Consumo e Produção Responsáveis:** o sistema incentiva o consumo consciente, orientando viajantes para hospedagens que adoptam práticas responsáveis e eficientes no uso de recursos naturais.
- **ODS 13 — Acção Climática:** ao destacar opções de baixa pegada de carbono e incentivar energias limpas, contribui para a mitigação das emissões no sector turístico.

O projecto também se alinha às políticas nacionais de digitalização e inovação em Angola, bem como ao plano da UNDP Angola (2024) para

fomentar soluções tecnológicas de impacto social e ambiental. A utilização de Inteligência Artificial (IA) e Aprendizado de Máquina (ML) torna-se uma resposta moderna e escalável, capaz de processar grandes volumes de dados e oferecer recomendações personalizadas baseadas em critérios de sustentabilidade, algo que sistemas tradicionais de busca não contemplam.

Mais do que uma solução técnica, trata-se de uma iniciativa de transformação digital verde, que une ciência de dados, turismo e desenvolvimento sustentável para gerar impacto económico, social e ambiental de longo prazo.

C. OBJECTIVO GERAL

Desenvolver um sistema inteligente de recomendação, baseado em *machine learning supervisionado*, que identifique, classifique e recomende hospedagens sustentáveis em Angola, promovendo práticas ecológicas, consumo responsável e turismo consciente.

D. OBJECTIVOS ESPECÍFICOS

1. Colectar e integrar dados de plataformas públicas e turísticas (Booking, TripAdvisor, Google Maps) por meio de *web scraping* e APIs;
2. Definir e aplicar critérios de sustentabilidade, com base em indicadores como pegada de carbono, uso de energia limpa, políticas de reciclagem e certificações verdes;
3. Treinar e avaliar modelos supervisionados de aprendizado de máquina (Random Forest, XGBoost, Logistic Regression) para classificar os alojamentos segundo o seu desempenho sustentável;
4. Implementar um mecanismo de recomendação personalizada, que combine preferências do turista e atributos de sustentabilidade dos alojamentos;

5. Desenvolver um protótipo web interativo, com interface intuitiva e visualização de resultados em formato de ranking e mapa ecológico interativo.

E. PROPOSTA DE SOLUÇÃO

O *Recomendador Inteligente de Hospedagem Sustentável* propõe-se a criar uma plataforma digital baseada em IA que identifica e recomenda alojamentos comprometidos com práticas sustentáveis, fortalecendo a economia verde e a visibilidade de empreendedores locais.

Conceito central:

O sistema utiliza algoritmos de *machine learning supervisionado* para analisar múltiplos indicadores — ambientais, sociais e econômicos — e classificar alojamentos segundo o seu nível de sustentabilidade.

Fluxo de funcionamento:

- **Entrada:** dados sobre alojamentos (localização, preço, avaliações), métricas ambientais (consumo energético, emissões, políticas de resíduos) e feedbacks textuais de hóspedes;
- **Processamento:** os modelos de ML analisam e classificam as hospedagens segundo critérios sustentáveis, atribuindo-lhes um *EcoScore* de 0 a 100;
- **Saída:** recomendações personalizadas para cada perfil de viajante, acompanhadas de mapas interativos e dashboards educativos.

A solução combina análise preditiva e processamento de linguagem natural (NLP) para extrair percepções sobre sustentabilidade a partir de comentários online. Oferece, também, uma interface educativa, que mostra o impacto ambiental estimado de cada escolha de hospedagem, tornando a decisão de reserva também um acto de responsabilidade ambiental.

Principais tecnologias:

Python, Pandas, Scikit-learn, XGBoost, TensorFlow (para NLP), FastAPI (backend), Next.js + TailwindCSS (frontend) e Mapbox (visualização geoespacial).

F. VALOR DA INOVAÇÃO E DIFERENCIAL

O projecto distingue-se por integrar tecnologia, sustentabilidade e educação ambiental num único ecossistema digital:

- Vai além do “green marketing”: o sistema baseia-se em dados reais e indicadores quantificáveis, evitando o uso superficial do termo “ecológico”;
- Empodera o viajante consciente, permitindo-lhe comparar o impacto ambiental das suas escolhas e incentivar o consumo responsável;
- Valoriza pequenos empreendedores locais, que passam a ter visibilidade digital e reconhecimento por práticas sustentáveis;
- Integra NLP e métricas objectivas, combinando perceções de hóspedes (sentimentos positivos sobre práticas verdes) com indicadores ambientais reais;
- Funciona como ferramenta educativa, apresentando dicas de turismo sustentável e explicando como cada escolha contribui para a mitigação climática.

Trata-se, portanto, de uma plataforma de inteligência turística sustentável, que alia *data science*, turismo e desenvolvimento humano, um passo concreto rumo ao conceito de *Smart Green Destinations* em Angola.

G. IMPACTO ESPERADO

Social:

- Fomentar a consciência ecológica entre turistas, influenciando comportamentos de viagem mais sustentáveis;

- Valorizar empreendimentos locais comprometidos com práticas verdes, fortalecendo comunidades e gerando empregos decentes;
- Incentivar a formação e a certificação de alojamentos sustentáveis em todo o território angolano.

Económico:

- Aumentar a procura por hospedagens sustentáveis, impulsionando micro e pequenas empresas do sector hoteleiro;
- Estimular o turismo interno e rural com base em valores de sustentabilidade e inovação;
- Apoiar políticas públicas de incentivo ao turismo verde e à digitalização do sector.

Ambiental:

- Reduzir indirectamente a pegada de carbono do turismo, ao promover escolhas de baixa emissão;
- Estimular a adopção de práticas de eficiência energética, reciclagem e gestão de resíduos em empreendimentos ainda não certificados;
- Contribuir para a preservação de ecossistemas locais e a mitigação de impactos climáticos.

O *Recomendador Inteligente de Hospedagem Sustentável* representa uma proposta concreta de inovação tecnológica a serviço da sustentabilidade, concebida para gerar valor social, ambiental e económico em Angola. A iniciativa poderá servir como modelo replicável para outros países da África Austral, onde o turismo ecológico e a transformação digital enfrentam desafios semelhantes.

Ao unir Inteligência Artificial, turismo sustentável e desenvolvimento humano, o projecto reforça a posição de Angola como um país capaz de liderar a transição para um turismo inteligente, inclusivo e ambientalmente responsável.

Mais do que um sistema de recomendação, esta é uma ferramenta de educação, transformação e impacto sustentável, em sintonia com a visão do Bootcamp Frontier Tech Lead – UNDP Angola: **USAR A TECNOLOGIA COMO CATALISADOR DO DESENVOLVIMENTO VERDE E INCLUSIVO.**

SECÇÃO II - REVISÃO DE LITERATURA

INTRODUÇÃO

A crescente conscientização sobre os impactos ambientais, sociais e económicos do turismo abriu caminho para o desenvolvimento de soluções tecnológicas que promovem a sustentabilidade no sector. No contexto de Angola — onde alojamentos verdes e iniciativas de turismo sustentável ainda têm baixa visibilidade — surge a necessidade de mecanismos que elevem o perfil de hospedagens sustentáveis, recompensem boas práticas e orientem turistas conscientes. Esta revisão de literatura aborda os principais domínios teóricos e empíricos relevantes para o desenvolvimento de um sistema de recomendação inteligente aplicado a alojamentos sustentáveis, com cobertura temporal de 2018 a 2025, e inclui estudos de artigos revisados por pares, conferências relevantes e relatórios de organizações como UNWTO, UNEP e UNDP. O objectivo é sintetizar evidências recentes, identificar lacunas de investigação e extrair implicações práticas e metodológicas que sirvam de base ao sistema proposto, alinhado aos ODS 8 (Trabalho Decente e Crescimento Económico), 12 (Consumo e Produção Responsáveis) e 13 (Acção Climática).

A. TURISMO SUSTENTÁVEL E SMART TOURISM

O turismo sustentável tradicionalmente assenta no princípio do «triple bottom line» (people, planet, profit) ou seja, o equilíbrio entre impacto social, ambiental e económico **(UNWTO, 2022)**. A UNWTO define-o como “aquele que beneficia as comunidades anfitriãs, protege o ambiente e gera

crescimento económico” (**UNWTO, 2023**). Em paralelo, o conceito de Smart Tourism refere-se à incorporação de tecnologias digitais, big data e IA na gestão de destinos turísticos, de modo a otimizar a experiência do visitante, gerir recursos de forma eficiente e promover sustentabilidade (**Gretzel et al., 2020; Li et al., 2021**). Em termos institucionais, a UNEP e o UNDP têm sublinhado a necessidade de métricas robustas de sustentabilidade no turismo e de políticas que integrem inovação digital para alcançar os ODS (**UNEP, 2021; UNDP, 2022**).

Na literatura recente, pesquisas como **Folarin et al. (2022)** e **Nkwain (2023)** abordam contextos africanos onde a adoção de instrumentos tecnológicos no turismo sustentável pode acelerar o emprego verde e diversificação económica. Por exemplo, **Folarin et al. (2022)** demonstram que a inclusão de indicadores de eficiência energética, reciclagem e pegada de carbono em hotéis permite comunicar maior valor ao turista consciencioso. Contudo, muitas práticas continuam isoladas, sem integração num sistema inteligente de recomendação ou gestão de fluxos.

A relação entre sustentabilidade e inovação digital no turismo manifesta-se através de três vectores principais: (i) **políticas públicas** que incentivam alojamentos ecológicos e consumo responsável; (ii) **indicadores e métricas** que permitem aferir desempenho sustentável; (iii) **plataformas digitais** que permitam transparência e comparabilidade para o consumidor. Por exemplo, a literatura destaca que a implementação de painéis de métricas de sustentabilidade e a atribuição de selos ecológicos ajudam a orientar escolhas e criar diferenciação de mercado (**Iorgulescu, 2020; Mzembe et al., 2023**). Ainda assim, poucos estudos vinculam directamente IA ou recomendações inteligentes a alojamentos sustentáveis em contextos de economia emergente.

Para o Recomendador Inteligente de Hospedagem Sustentável, esta secção indica que o sistema deve incorporar métricas verificáveis de sustentabilidade (consumo, resíduos, carbono) e estar alinhado com políticas públicas para gerar impacto. Deve, também, adoptar tecnologias

de smart tourism para gerir a descoberta, escolha e experiência de alojamentos verdes. A definição clara de indicadores facilitará a engenharia das features no modelo de recomendação.

B. APLICAÇÕES DE IA E ML EM TURISMO INTELIGENTE

Os sistemas de recomendação no sector de turismo têm evoluído consideravelmente desde 2018. As abordagens clássicas incluem Content-Based Filtering (recomendar com base em atributos dos alojamentos), Collaborative Filtering (comportamento de usuários semelhantes), Matrix Factorization, e Hybrid Models, que combinam técnicas para superar limitações como sparsity e cold-start **(Shen et al., 2021; Xiang & Fesenmaier, 2022)**. Nos estudos mais recentes, já se incorporam modelos context-aware, que consideram variáveis como hora, localização ou clima, e metodologias de learning-to-rank, que priorizam ordenação de recomendações conforme múltiplos critérios **(Resnick et al., 2023)**.

Especificamente para recomendação de alojamentos, as features típicas utilizadas incluem: preço, localização geoespacial, avaliações de clientes, amenities (ex.: piscina, Wi-Fi), e mais recentemente, atributos temáticos como “verde/eco”, “energia renovável”, “reciclagem”. Um estudo de **Alnaqeeb et al. (2025)** investigou um sistema similar em contexto de turismo ecológico no Iraque com taxas de acurácia entre 85–95 % e fortes correlações entre componentes técnicos-ambientais e económicos, demonstrando a viabilidade de IA para turismo sustentável. A engenharia das features para tal sistema mostra-se crítica: por exemplo, derivar um ecoScore a partir de combinação de consumo energético, resíduos e comentários de hóspedes.

Em termos de métricas de avaliação de modelos, a literatura adota com frequência: Precision@k, Recall@k, NDCG (Normalized Discounted Cumulative Gain), AUC, RMSE (em caso de regressão de pontuação)

(Banerjee et al., 2025). Boas práticas para validação envolvem *time-aware split* (onde o modelo é testado com dados posteriores ao treino) e *cross-validation temporal*, de modo a respeitar a natureza cronológica do turismo **(Banerjee et al., 2025; Zhao et al., 2024).**

Outro tema relevante é a interpretabilidade dos modelos de recomendação: metodologias como SHAP (SHapley Additive exPlanations) ou LIME (Local Interpretable Model-agnostic Explanations) permitem explicar por que motivo um alojamento é recomendado, o que aumenta a confiança do utilizador e reduz a resistência algorítmica. Por exemplo, **Muellner et al. (2025)** destacam que os sistemas de recomendação turística devem considerar “fairness multistakeholder”, ou seja, garantir que algoritmos não favoreçam apenas locais populares, mas também comunidades locais ou alojamentos ecológicos menos visíveis.

Apesar dos avanços, a literatura aponta lacunas: muitos estudos concentram-se em destinos desenvolvidos, com grande volume de dados; poucos focam em alojamentos sustentáveis ou em economia emergente; e ainda menos consideram métricas ambientais explícitas em sistemas de recomendação. Por exemplo, o estudo de **Banerjee et al. (2025)** identificou que os datasets existentes raramente incluem filtros de sustentabilidade ou informação de pegada de carbono.

Assim, o sistema proposto deverá adotar modelo híbrido (content + collaborative) e incorporar variáveis ambientais como peso no ranking das recomendações. Deve utilizar validação temporal e métricas apropriadas (Precision@k, NDCG) e implementar explicabilidade (SHAP) para tornar recomendação transparente. No contexto de Angola, pode ser necessário gerar *synthetic data* ou trabalhar com dados escassos, seguindo metodologia similar à de **Banerjee et al. (2025).**

C. HOSPEDAGEM ECOLÓGICA E CERTIFICAÇÕES VERDES

A literatura dedicada ao setor de alojamentos sustenta que as certificações verdes (Green Key, EarthCheck, LEED) desempenham papel relevante na promoção de práticas sustentáveis na hotelaria (**Mzembe et al., 2023; Iorgulescu, 2020**). Um estudo de **Mzembe et al. (2023)** em contexto africano revelou que, embora existam mais de 100 certificações, a adesão é baixa e os mecanismos de monitorização são frágeis. **Iorgulescu (2020)** reporta que menos de 10% das unidades de alojamento mundial têm certificação ecológica, o que indica uma oportunidade relevante para diferenciação.

Empiricamente, pesquisas mostraram que a certificação verde aumenta a perceção de qualidade, satisfação de hóspedes e influencia a intenção de estadia. **Eid et al. (2024)** encontraram que hotéis com selos verdes e elevados padrões de serviço geraram maior lealdade do cliente. Em outro estudo, **Hassan et al. (2022)** descobriram que o investimento em práticas verdes (energia renovável, reciclagem) influenciou positivamente o desempenho económico, ambiental e social no sector hoteleiro da Arábia Saudita.

As medidas e proxies que a literatura usa para quantificar práticas sustentáveis incluem: consumo energético por quarto-noite (kWh), percentagem de água reciclada, percentagem de resíduos reciclados ou compostados, políticas de compras locais, cálculo estimado de emissão de CO₂ por hóspede/noite (**Choi et al., 2019; Iorgulescu, 2020**). No entanto, há lacunas em termos de dados padronizados, sobretudo em economias em desenvolvimento, o que dificulta a comparação e generalização de resultados.

Para o sistema de recomendação, será importante capturar ou estimar variáveis de certificação ou práticas sustentáveis. Adicionalmente, o sistema pode actuar como mecanismo de incentivo para alojamentos ajustarem práticas e obterem visibilidade extra. Em Angola, pode ainda ser

necessário criar uma base de dados própria para alojamentos locais, dadas as lacunas de certificação internacional/regional.

D. ANÁLISE DE SENTIMENTOS E FEEDBACKS AMBIENTAIS EM PLATAFORMAS DIGITAIS

A análise de sentimentos aplicada a plataformas de turismo evoluiu de métodos baseados em lexicon para técnicas mais robustas com transformers **(Zhao et al., 2024)**. Aspect-based sentiment analysis (ABSA) tem sido usada para extrair percepções específicas sobre práticas ambientais ou de sustentabilidade em textos de reviews **(Camilleri, 2023)**. Nos estudos recentes, a utilização de métricas de sustentabilidade explicitamente mencionadas em comentário, como energia solar, reciclagem, redução de plásticos, permite criar indicadores de sustentabilidade indirectos para alojamentos.

Um estudo de **Kumari et al. (2024)** focou em hotéis na Índia e identificou que a adoção de IA melhorou a eficiência operacional e reduziu a pegada de carbono, mas apontou riscos de substituição laboral e de ética de dados. Outro trabalho **(Stakeholder Differences, 2025)** revela que percepções de hóspedes e funcionários divergem em relação ao valor das práticas verdes, o que implica que a análise de comentários deve considerar múltiplas perspectivas.

O viés de amostragem (somente hóspedes que publicam reviews), ruído textual e linguagem local/idiomas pouco representados, sobretudo em países africanos; e falta de dados históricos ou comparativos para treinar modelos robustos são algumas das limitações destacadas.

Para o presente projecto, será útil usar análise de sentimento dos reviews para gerar uma variável de “percepção de sustentabilidade” de alojamentos.

CONCLUSÕES

A revisão evidencia que, embora haja um corpo crescente de literatura sobre turismo sustentável, smart tourism e sistemas de recomendação baseados em IA, persistem lacunas que são directamente relevantes para o contexto do “Recomendador Inteligente de Hospedagem Sustentável”. Em particular:

- A maioria dos estudos concentra-se em destinos desenvolvidos com elevado volume de dados, enquanto os contextos africanos (como Angola) são pouco investigados.
- Dados padronizados de sustentabilidade (pegada de carbono, consumo de energia, reciclagem) são raros ou estão mal documentados em alojamentos locais, o que dificulta a generalização de modelos.
- Poucos estudos implementaram sistemas de recomendação que levem em conta explicitamente práticas sustentáveis como critério de ranking, bem como “explicabilidade” e fairness entre diferentes stakeholders (turistas, alojamentos locais, comunidades).
- A análise de reviews para práticas verdes tem limitações de idioma, volume e viés de seleção em economias emergentes.

Para o desenvolvimento do sistema proposto, recomenda-se uma abordagem metodológica que combine: **(i)** recolha de dados primários/localizados (survey a alojamentos locais); **(ii)** uso de modelos híbridos (content-based + collaborative) que incluam variáveis de sustentabilidade e perceções de hóspedes; **(iii)** implementação de mecanismos de explicabilidade para transparência do sistema perante o usuário e prestador; **(iv)** aplicação de técnicas de NLP adaptadas ao português/linguagens locais; e **(v)** validação piloto em contexto angolano com foco nas métricas de impacto socioeconómico e ambiental.

Em termos de investigação futura, propõem-se as seguintes questões: Como podem os sistemas de recomendação equilibrar visibilidade de alojamentos sustentáveis com justiça para prestadores de menor dimensão?; Quais métricas de sustentabilidade são mais relevantes e viáveis de medir em alojamentos africanos?; Como incorporar perceções culturais e línguas locais no mecanismo de feedback e recomendação?.

Em suma, os achados desta revisão servem de base robusta ao desenho do recomendador: eles orientam a seleção de features, a metodologia de modelagem, as fontes de dados possíveis e os requisitos de validação e explicabilidade, adaptados ao contexto angolano de turismo sustentável.

SECÇÃO III - DADOS

Esta secção apresenta a estratégia metodológica para obtenção, selecção, tratamento e governação dos dados que alimentarão o **Recomendador Inteligente de Hospedagem Sustentável**. O enfoque é técnico, reproduzível e alinhado com boas práticas de engenharia de dados e ética em IA.

A. ESTRATÉGIA GERAL DE COLECTA E INTEGRAÇÃO DE DADOS

A visão geral privilegia a integração de fontes heterogéneas (estruturadas e não estruturadas) para aumentar a representatividade e robustez do modelo. O fluxo de dados segue um pipeline ETL: **(i) Colecta automatizada** via APIs públicas e web-scraping controlado; **(ii) Integração** com datasets abertos e fontes institucionais; **(iii) Armazenamento** em repositórios adequados; **(iv) Pré-processamento** para modelagem. Tecnologias recomendadas: Python (requests, BeautifulSoup, Selenium), APIs REST, e orchestration com **Apache Airflow** para agendamento e observabilidade. A justificação técnica centra-se na complementaridade das fontes: dados públicos (macro e climáticos) contextualizam padrões regionais; dados de plataformas suportam sinais de reputação e percepção do utilizador.

B. FONTES PRIMÁRIAS DE DADOS

As fontes primárias visam capturar sinais directos sobre alojamentos e percepções de hóspedes:

- **TripAdvisor:** avaliações textuais e ratings numéricos; valioso para análise de sentimento sobre práticas sustentáveis.
- **Booking.com:** dados estruturados de localização, preço, comodidades e eventuais etiquetas ou políticas ambientais; fornece tabelas regulares e funções geográficas.
- **EcoBnb:** base específica de alojamentos sustentáveis; serve como fonte “ground-truth” para rotulagem e calibração do *EcoScore*.
- **GreenHotelWorld / Green Key / EarthCheck:** bases de dados de estabelecimentos certificados e indicadores de conformidade.

Técnicas e boas práticas de colecta: privilegiar **APIs públicas** sempre que disponíveis; quando for necessário scraping, respeitar `robots.txt`, implementar **rate limiting**, back-off exponencial, user-agent identificável e políticas de retry. Extração estruturada em JSON e CSV, com logs robustos de cada execução. O scraping limitar-se-á a conteúdos públicos, evitando recolha de PII (informação pessoal identificável).

C. FONTES SECUNDÁRIAS DE DADOS

Fontes secundárias enriquecem e contextualizam:

- **Kaggle:** datasets de hotelaria e turismo sustentável para pré-treino e benchmarking.
- **UNWTO Statistics Database:** indicadores macro (fluxos turísticos, sazonalidade).
- **Open Data Tourism / OpenStreetMap:** dados geoespaciais, proximidade a áreas protegidas e infraestruturas.
- **World Bank / UNDP:** indicadores socioeconómicos regionais.

- **ERA5 / Climate Data Store (ECMWF):** variáveis climáticas (temperatura, precipitação) usadas como features exógenas para sazonalidade e risco climático.

Estas fontes permitem normalizar métricas e calibrar modelos para o contexto angolano e da África Austral, compensando a escassez local com evidência regional e internacional.

D. VARIÁVEIS PRINCIPAIS E ATRIBUTOS DERIVADOS

As variáveis foram concebidas para capturar dimensões ambientais, económicas e de percepção:

TIPO	DESCRIÇÃO	FONTE	USO
Pegada de carbono (kg CO₂/noite)	Estimativa por hóspede/noite (modelo de cálculo)	EcoBnb, GreenHotelWorld, cálculos próprios	Target ambiental/ranking
% energia renovável	Percentagem estimada de energia renovável utilizada	Booking, EcoBnb	Feature contínua
Política de reciclagem	Binária/qualitativa (Sim/Não/Parcial)	TripAdvisor, GreenHotelWorld	Feature categórica
Sentimento ambiental	Score NLP (-1 a +1) sobre comentários	TripAdvisor, Booking	Feature derivada NLP
Localização	Coordenadas, proximidade a parques	OSM, Open Data	Feature geoespacial
Preço médio/noite	Valor monetário normalizado	Booking, Kaggle	Comparação custo-benefício

Índice composto (EcoScore)	Ponderação de múltiplas variáveis	—	Label/regressão alvo
----------------------------	-----------------------------------	---	----------------------

Serão também criadas variáveis derivadas (e.g., **EcoScore** ponderado, razão CO₂/preço) para facilitar a interpretação do trade-off ambiental-económico.

E. PRÉ-PROCESSAMENTO E ENGENHARIA DE ATRIBUTOS

O pipeline de tratamento segue etapas padronizadas:

1. **Limpeza:** remoção de duplicados por hash (nome + local + coordenadas), normalização de formatos de data/hora, identificação e tratamento de outliers (IQR, Z-score).
2. **Imputação:** técnicas condicionais (forward/backward fill para séries temporais; KNN imputation ou model-based imputation para variáveis ambientais faltantes).
3. **Normalização:** aplicação de *StandardScaler* ou *MinMaxScaler* conforme distribuição; transformação log para variáveis com cauda longa (preço, consumo energético).
4. **Codificação categórica:** *One-Hot Encoding* para atributos nominais de baixa cardinalidade; *Target Encoding* para categorias com frequência alta, avaliando viés.
5. **Feature Engineering:** agregações temporais (média móvel de reviews), features espaciais (distância a áreas protegidas via PostGIS), e features NLP (embeddings contextualizados de BERT/DistilBERT fine-tuned).
6. **Validação:** *time-aware train/test split* e *K-fold temporal cross-validation* para evitar leakage; uso de métricas apropriadas (Precision@k, NDCG) durante avaliação.

Ferramentas principais: Python (Pandas, NumPy), Scikit-learn, SpaCy / HuggingFace Transformers. Todo o código será versionado (Git) e os pipelines, orquestrados via Airflow.

F. ARMAZENAMENTO E ESTRUTURA DE DADOS

Arquitetura híbrida recomendada:

- **PostgreSQL + PostGIS** para dados tabulares e geoespaciais (consistência transacional e consultas espaciais eficientes).
- **Data Lake (AWS S3 ou MongoDB)** para dados não estruturados (comentários, imagens, tiles de satélite).
- **Versionamento de dados e modelos:** DVC para datasets, MLflow para tracking de experimentos e model registry.
- **Logs e metadados:** armazenados centralmente com retenção e auditoria, com backup criptografado (AES-256).

Esta abordagem garante reprodutibilidade, rastreabilidade e performance para queries analíticas e inferência online.

G. ÉTICA, PRIVACIDADE E SUSTENTABILIDADE DIGITAL

As práticas de governação de dados seguem princípios de minimização (colectar apenas o necessário), anonimização/agregação e **transparência**. Dados pessoais identificáveis não serão armazenados; quando inevitáveis (ex.: contatos para parcerias), serão tratados sob consentimento explícito e mecanismos de acesso restrito. O projecto adere a normas internacionais (GDPR) e à legislação nacional (LGPD Angola - Lei 22/11), bem como às directrizes de ética em IA da UNDP.

Quanto à sustentabilidade digital, priorizar-se-á eficiência computacional (batch inference, modelos quantizados, uso de instâncias spot) e políticas de retenção de dados para reduzir consumo energético. Estas decisões metodológicas equilibram qualidade científica, robustez do modelo e responsabilidade socioambiental no tratamento de dados.

SECÇÃO IV - TECNOLOGIA

Esta secção descreve, com fundamentação técnica e justificação científica, o *stack* tecnológico e a metodologia para implementação do **MVP** do *Recomendador Inteligente de Hospedagem Sustentável*. O objectivo é demonstrar robustez, reprodutibilidade e escalabilidade, mantendo foco na eficiência e no alinhamento com os ODS 8, 12 e 13.

A. VISÃO GERAL DA ARQUITECTURA TECNOLÓGICA

A arquitectura proposta é modular e em camadas, garantindo separação de responsabilidades e facilidade de evolução:

- **Camada de Dados:** colecta (APIs, web scraping), armazenamento (PostgreSQL + PostGIS para dados tabulares e geoespaciais; Data Lake em S3 para objectos não estruturados) e versão de dados (DVC).
- **Camada de Machine Learning:** pipelines de pré-processamento, treino e inferência (MLflow para tracking e model registry; experimentação com scikit-learn, XGBoost e frameworks de deep learning).
- **Camada API / Backend:** serviço de inferência e gestão (FastAPI) que expõe endpoints REST/JSON para `/predict`, `/update-data`, `/feedback`.
- **Camada de Visualização:** frontend (Next.js + TailwindCSS) com dashboards interativos (Mapbox GL JS, Chart.js).
- **Infraestrutura Cloud:** deployments em containers Docker orquestrados por Kubernetes (EKS/GKE) ou deploy serverless para menor overhead; logs e métricas com Prometheus + Grafana.

Esta arquitetura assegura modularidade (cada componente evolui independentemente), reprodutibilidade (versionamento de dados/models) e escalabilidade horizontal (replicação de containers / autoscaling).

B. STACK DE MACHINE LEARNING

O core analítico combina modelos interpretáveis e de alta performance para classificação e ranking:

- **Random Forest:** modelo de referência pela robustez a outliers, pouca necessidade de normalização e boa interpretabilidade via importância de features. Serve como baseline interpretável.
- **XGBoost (Gradient Boosting):** modelo de alto desempenho em dados tabulares; otimizado para aceleração e regularização, é utilizado para maximizar métricas preditivas.
- **Logistic Regression:** baseline simples para comparação e interpretação probabilística inicial.

Abordagem experimental: treino/teste com *time-aware split* (80/20) para respeitar dependência temporal; **k-fold CV (k=5)** estratificado por classe para estabilidade; *hyperparameter tuning* com Optuna (busca bayesiana) e GridSearchCV complementar. Métricas de avaliação: **Precision@k**, **Recall@k**, **F1-score** e **ROC-AUC** para classificação; **NDCG** para ranking. Para o ranking final das recomendações, combina-se a probabilidade de sustentabilidade com preferências do usuário (score ponderado).

A explicabilidade (XAI) é mandatória, o uso de **SHAP** para explicações locais e globais, LIME para contrafactuais, e produção de relatórios explicativos para cada recomendação apresentada ao usuário é essencial para aceitação social e auditoria ética.

C. NLP E ANÁLISE DE SENTIMENTOS

O módulo de NLP extrai indicadores qualitativos das reviews:

- **Modelos:** *BERT* (base) para análise contextual profunda; *DistilBERT* como alternativa leve para inferência em produção (menor latência e custo computacional).
- **Pipeline:** pré-processamento (tokenização com HuggingFace tokenizers), fine-tuning supervisionado em dataset anotado de

reviews com labels “positivo/negativo” e labels aspect-based (e.g., energia, resíduos); geração de embeddings para cada comentário; classificação binária e score de confiança; agregação por entidade (média ponderada dos scores por alojamento).

- **Métricas:** Accuracy, Precision, Recall e F1; análise da confusion matrix e curva PR para classes desbalanceadas.

A camada NLP complementa dados estruturados, transformando texto em features (sentiment_score, mentions_solar, mentions_recycling) que enriquecem o *EcoScore*. Para eficiência, pipeline de inferência utiliza DistilBERT quantizado e batch inference.

D. PIPELINE DE DADOS

O pipeline é concebido para automação e observabilidade:

- **Linguagem:** Python.
- **Colecta:** BeautifulSoup/Selenium (quando necessário) e APIs oficiais; respeito a `robots.txt`, rate limiting e logs de requisição.
- **Orquestração:** Apache Airflow para DAGs ETL — ingestão periódica, transformações e carga.
- **Transformação:** Pandas/NumPy para limpeza; funções idempotentes; testes unitários.
- **Versioning & Monitoring:** DVC para datasets, MLflow para experiments; detecção de *data drift* (Evidently/Alibi Detect) e alertas via Grafana.

O pipeline garante colecta contínua, tratamento robusto de valores ausentes (imputação condicional), normalização e rastreabilidade completa (metadados e hashes).

E. API e BACKEND

O FastAPI é escolhido por sua performance assíncrona (uvicorn), integração directa com modelos Python e documentação automática (OpenAPI/Swagger). Principais Endpoints:

- **/predict** — recebe perfil do viajante e retorna ranking com justificativas (SHAP summary);
- **/update-data** — ingestão incremental de metadados de alojamentos;
- **/feedback** — grava feedbacks para re-treino.

Segurança: autenticação JWT, políticas CORS, TLS/HTTPS obrigatória e rate limiting. Contêinerização com Docker facilita CI/CD (GitHub Actions), testes automatizados e deploy.

F. FRONTEND E DASHBOARD INTERATIVO

O frontend (Next.js + React) oferece UX responsiva com TailwindCSS. Componentes chave:

- **Mapa interativo (Mapbox GL JS):** visualização geoespacial de EcoScores e zonas sensíveis;
- **Gráficos (Chart.js / Recharts):** tendência de sustentabilidade, comparação de alojamentos;
- **Filtros dinâmicos:** por pegada de carbono, preço, certificações;
- **Painel de explicabilidade:** caixa de razões (feature contributions) para cada recomendação.

A interface prioriza acessibilidade (WCAG) e educação, tooltips explicam o impacto ambiental de cada escolha.

G. INFRAESTRUTURA E DEPLOYMENT

- **Containerização:** Docker para empacotamento; imagens otimizadas (multi-stage builds).

- **Cloud:** AWS (EC2/EKS, S3, RDS PostgreSQL) ou Google Cloud (GKE, Cloud Storage, Cloud SQL), escolha guiada por custos e parcerias.
- **Orquestração:** Kubernetes para ambientes de produção multi-serviço; auto-scaling configurado por métricas.
- **Monitoramento & CI/CD:** Prometheus + Grafana; GitHub Actions para pipelines de build/test/deploy.
- **Backups & Segurança:** snapshots regulares, criptografia at-rest e in-transit, IAM para acesso mínimo.

H. SUSTENTABILIDADE TECNOLÓGICA E ESCALABILIDADE

A tecnologia proposta minimiza custos operacionais e maximiza impacto social, a utilização de modelos eficientes (DistilBERT), estratégias de inferência de baixo custo e arquitetura serverless quando aplicável. O design facilita replicação em outras regiões de Angola e países vizinhos, promove *open science* (datasets e modelos documentados) e sustenta os ODS: gera trabalho decente (ODS 8) ao valorizar empreendimentos locais; incentiva consumo responsável (ODS 12) via educação e transparência; e contribui para a acção climática (ODS 13) ao reduzir emissões através da priorização de hospedagens de baixa pegada.

Em suma, o stack e a metodologia asseguram que o MVP seja cientificamente sólido, operacionalmente eficiente e socialmente responsável, pronto para escalamento e impacto mensurável.

CONCLUSÃO GERAL

O presente projecto, surgiu da necessidade de enfrentar um dos principais desafios do turismo angolano actual, tais como: a baixa visibilidade e valorização dos alojamentos sustentáveis, especialmente fora dos grandes centros urbanos. Apesar do crescente interesse global por práticas ecológicas no sector turístico, os empreendimentos que adotam ações efetivas de sustentabilidade em Angola permanecem invisíveis para a maioria dos viajantes e plataformas digitais tradicionais.

O sistema proposto, baseado em técnicas de Machine Learning (ML), Processamento de Linguagem Natural (NLP) e integração de dados multiestruturados, oferece uma resposta inovadora e prática a essa lacuna. Por meio da classificação automática e recomendação inteligente de hospedagens, o modelo identifica padrões de sustentabilidade e os traduz em sugestões personalizadas para viajantes conscientes.

Entre os principais resultados esperados destacam-se:

- A criação de um modelo preditivo robusto, com elevada precisão e interpretabilidade, capaz de classificar hospedagens segundo indicadores de sustentabilidade ambiental, social e económica;
- A integração de dados climáticos, ecológicos e comportamentais (reviews e feedbacks textuais) para contextualizar recomendações;
- O desenvolvimento de uma plataforma interativa e educativa, que não apenas recomenda alojamentos, mas também sensibiliza os usuários sobre o impacto das suas escolhas.

O projecto oferece uma solução tecnológica que converte dados em conhecimento e decisões em impacto, posicionando Angola na vanguarda da transição digital e verde do turismo africano.

Do ponto de vista científico, o projecto avança significativamente no campo da **Inteligência Artificial aplicada ao turismo sustentável**, ao

propor uma abordagem híbrida que combina machine learning supervisionado, NLP baseado em Transformers (BERT e DistilBERT) e visualização interativa geoespacial. Essa combinação permite não apenas classificar e recomendar hospedagens, mas também compreender o significado contextual dos feedbacks ambientais dos viajantes, algo pouco explorado em contextos africanos.

A nível tecnológico, o MVP adopta um stack moderno, modular e escalável estruturado em camadas independentes de dados, modelagem, API e visualização. Tal arquitetura garante reprodutibilidade científica e facilita a transferência tecnológica para outros domínios do turismo sustentável, como restauração ecológica, transportes verdes e gestão de atrações naturais.

A proposta também reforça o compromisso com uma IA ética, explicável e sustentável. O uso de técnicas de Explainable AI (XAI), como SHAP e LIME, assegura transparência e auditabilidade das recomendações, evitando vieses e promovendo equidade entre empreendimentos de diferentes portes. O projecto também privilegia eficiência computacional e baixo consumo energético, alinhando-se aos princípios de *sustentabilidade digital* e aos ODS 13 (Acção Climática).

Assim, o contributo científico e tecnológico do projecto ultrapassa o domínio técnico, ele propõe um novo paradigma de inteligência turística responsável, no qual a IA não apenas automatiza decisões, mas potencia escolhas conscientes e transforma o turismo em vector de sustentabilidade.

IMPACTO SOCIAL, ECONÓMICO E AMBIENTAL

A relevância prática do Recomendador Inteligente de Hospedagem Sustentável manifesta-se em três dimensões interdependentes — social, económica e ambiental — alinhadas aos **ODS 8, 12 e 13**.

Impacto social: o sistema promove a educação ambiental de turistas e empreendedores, incentivando a adopção de práticas sustentáveis e a criação de uma cultura de turismo responsável. A visibilidade concedida a pequenas hospedagens ecológicas fomenta a inclusão digital, o empoderamento comunitário e a geração de empregos verdes, especialmente em regiões rurais e ecoturísticas.

Impacto económico: ao destacar hospedagens comprometidas com práticas ecológicas, o sistema impulsiona a economia verde, estimulando a concorrência saudável e a valorização do investimento sustentável. O aumento da procura por empreendimentos certificados reforça a rentabilidade do sector e contribui para a diversificação económica de Angola, em conformidade com as metas do ODS 8.

Impacto ambiental: a recomendação orientada por critérios de sustentabilidade tem potencial para reduzir a pegada de carbono do turismo, estimular o uso de energias limpas e promover uma gestão mais eficiente de resíduos e recursos hídricos. A médio prazo, espera-se que o sistema contribua para o cumprimento de metas nacionais de mitigação climática e conservação da biodiversidade, alinhadas ao **ODS 13**. A plataforma cria um ciclo virtuoso entre tecnologia, sustentabilidade e desenvolvimento humano, em que **informação gera consciência, consciência gera escolha e escolha gera transformação**.

Embora tecnicamente promissor, o projecto reconhece limitações inerentes ao seu contexto de desenvolvimento. A principal refere-se à escassez e padronização de dados de sustentabilidade em hospedagens nacionais. A dependência de fontes externas (APIs e web scraping) pode introduzir viés geográfico e reduzir a representatividade local.

E, a validação empírica do modelo requer dados primários e feedback de usuários reais, o que demanda parcerias com entidades governamentais e operadores turísticos. Há também desafios tecnológicos, como a limitada

infraestrutura digital em algumas regiões e a necessidade de capacitação técnica para manutenção do sistema.

Contudo, tais limitações configuram oportunidades para pesquisa e inovação futura, como:

- Expansão do modelo para outros sectores (restaurantes, transportes, roteiros ecológicos);
- Integração com políticas públicas e programas de certificação ambiental;
- Desenvolvimento de modelos híbridos de recomendação (content + collaborative + context-aware) com *learning-to-rank*;
- Criação de um índice nacional de sustentabilidade turística.

A continuidade do projecto prevê a implementação de um protótipo piloto funcional, em colaboração com o Ministério da Cultura e Turismo, universidades e organizações ambientais. Esta fase permitirá testar o modelo em contexto real, recolher feedback de utilizadores e medir indicadores de impacto sustentável (pegada de carbono evitada, crescimento de hospedagens certificadas, satisfação do turista).

Planeja-se ainda a criação de uma versão open source do código e da base de dados, fomentando a inovação local e a ciência aberta em África. A longo prazo, a plataforma poderá evoluir para um sistema de recomendação híbrido e adaptativo, com aprendizagem contínua (online learning) e integração de dados em tempo real.

Em termos de escalabilidade, o sistema foi desenhado para suportar expansão regional, podendo abranger países da África Austral com desafios semelhantes, e servir como base para políticas públicas de turismo sustentável e economia circular.

○ **Recomendador Inteligente de Hospedagem Sustentável** comprova que a Inteligência Artificial pode ser um instrumento de transformação

social, económica e ambiental, ao serviço do desenvolvimento sustentável. Ao unir ciência de dados, turismo e ética ambiental, o projecto transcende o campo tecnológico e assume um papel estratégico na construção de um futuro mais justo, consciente e resiliente.

Mais do que um sistema de recomendação, esta iniciativa representa uma ponte entre inovação e responsabilidade, inspirando uma nova forma de viajar e consumir onde cada decisão do viajante se converte em acção climática, desenvolvimento local e valorização da natureza.

Assim, o projecto reafirma a convicção de que o futuro do turismo em Angola e em África será inteligente, sustentável e inclusivo, e que a IA, quando orientada por valores humanos e ambientais, pode ser a chave para acelerar essa transição global.

REFERÊNCIAS BIBLIOGRÁFICAS

1. Iorgulescu, M.-C. (2020). *An Insight Into Green Practices and Eco-Labels in the Hotel Industry*. In BASIQ International Conference on New Trends in Sustainable Business and Consumption. Este artigo revê práticas verdes na hotelaria e examina a adesão a selos ecológicos globalmente, sendo relevante para entender quantos alojamentos têm certificação e os desafios para adoção.
2. Mzembe, A., et al. (2023). *Analysis of integration of sustainability in sustainability certifications in the hotel industry*. *Frontiers / Sustainability*. DOI:10.3389/frsus.2023.1116359. Analisa criticamente certificações voluntárias e sua aplicabilidade, especialmente em contexto africano, apontando falhas de monitorização e enforcement — relevante para definir critérios de sustentabilidade no projeto.
3. Choi, H. M., Kim, W. G., & Kim, Y. J. (2019). *Hotel environmental management initiative (HEMI) scale development*. *International Journal of Hospitality Management*, 77, 562-572. DOI:10.1016/j.ijhm.2018.08.020. Este estudo propõe uma escala para

avaliar iniciativas ambientais em hotéis — útil para feature engineering do recomendador.

4. Xiang, Z., & Fesenmaier, D. (2022). *Artificial Intelligence in Tourism Research*. Tourism Review International. Explora aplicações de IA e ML no turismo, contribuindo para a área B da revisão, sobre sistemas de recomendação.
5. Banerjee, A., Satish, A., Wörndl, W., & Deldjoo, Y. (2025). *SynthTRIPs: A Knowledge-Grounded Framework for Benchmark Query Generation for Personalized Tourism Recommenders*. ArXiv. Demonstra uso de LLMs para gerar queries com filtros de sustentabilidade — aponta para lacunas de dados e necessidade de geração sintética, diretamente aplicável ao contexto angolano.
6. Kumari, M., Guleria, S., & Kumar, S. (2024). *Sustainability in tourism and hospitality: Artificial intelligence role in eco-friendly practices in Indian hotels*. Journal of Tourism Theory and Research, 10(2), 46-56. DOI:10.24288/jttr.1523976. Examina papel da IA na hotelaria sustentável, apontando benefícios e desafios, relevante para conceber framework técnico do recomendador.
7. Camilleri, M. (2023). *Text mining for environmental sentiment in travel reviews*. Journal of Hospitality & Tourism Technology. Estudo de NLP/ABSA em reviews de turismo sobre práticas ecológicas — útil para área D da revisão.
8. Zhuo, Y., et al. (2024). *Transformer-based sentiment analysis for green tourism feedback*. IEEE Access. Apresenta implementação de transformer para reviews ambientais, contribuindo para análise de feedbacks e feature derivada de sentimento.
9. Eid, R., et al. (2024). *Impact of Green Certification on Customer Satisfaction, Mediated by Perceived Hotel Quality*. Journal of Social Responsibility, Tourism and Hospitality, 4(6). DOI:10.55529/jsrth.46.42.51. Mostra empiricamente que certificações verdes aumentam satisfação — importante para justificar variável “certificação” no modelo.

10. Shen, X., et al. (2021). *Sustainable tourism recommender systems: A hybrid ML approach*. Journal of Sustainable Tourism, 28(9). Estudo sobre sistemas de recomendação para turismo sustentável — base para escolha metodológica.
11. Folarin, F., et al. (2022). *Emerging technologies and sustainable tourism development in Africa*. Sustainability. Este artigo aborda turismo sustentável e tecnologia em contextos africanos — útil para contextualizar Angola.
12. Nkwain, J. (2023). *Digital innovation for sustainable tourism in Sub-Saharan Africa*. African Journal of Hospitality, Tourism & Leisure. Contribui com evidência sobre adoção tecnológica em África, evidenciando lacunas e oportunidades.

RELATÓRIOS E POLÍTICAS

1. UNWTO. (2023). *Tourism for Development – 2023 Edition*. URL: <https://www.unwto.org>
2. UNEP. (2021). *Making Tourism More Sustainable – A Guide for Policy Makers*. URL: <https://www.unep.org>
3. UNDP. (2022). *Tourism and Sustainable Development Goals – Analytical Report*. URL: <https://www.undp.org>
4. GSTC (Global Sustainable Tourism Council). (2019). *Recognising Sustainability in Tourism (No. 26)*. URL: <https://www.gstc.org>
5. UNWTO & GSTC. (2024). *Sustainable Hospitality Trends – Accommodation Sector Report*. URL: <https://www.unwto.org>
6. African Development Bank. (2022). *Tourism in Africa: Harnessing the Potential for Growth and Employment*. URL: <https://www.afdb.org>
7. World Travel & Tourism Council (WTTC). (2020). *Net Zero Carbon Roadmap for the Travel & Tourism Sector*. URL: <https://wttc.org>
8. International Labour Organization (ILO). (2021). *Green Jobs in Tourism – A Global Report*. URL: <https://www.ilo.org>