

## DESENVOLVIMENTO DE UM MODELO DE RECOMENDAÇÃO PARA A PERSONALIZAÇÃO DA EXPERIÊNCIA TURÍSTICA

MANUELA OTAVIO DA SILVA<sup>1</sup>, MARIO TADASHI SHIMANUKI<sup>2</sup>

<sup>1</sup>Graduanda em Tecnologia em Análise e Desenvolvimento de Sistemas, IFSP, Câmpus Caraguatatuba, o.manuela@aluno.ifsp.edu.br.

<sup>2</sup> Professor no Instituto Federal de Educação, Ciência e Tecnologia de São Paulo, IFSP, Câmpus Caraguatatuba, mario@ifspcaragua.net.

**Área de conhecimento (Tabela CNPq):** Sistemas de Informação – 1.03.03.05-7

**RESUMO:** O presente trabalho apresenta o desenvolvimento de um modelo híbrido de recomendação para personalizar a experiência turística. Considerando o paradoxo da escolha e a sobrecarga de informações enfrentada pelos turistas, o modelo utiliza redes neurais profundas e a biblioteca TensorFlow Recommenders (TFRS) em uma arquitetura de duas etapas: recuperação e classificação. A abordagem incorpora embeddings e uma camada LSTM (Long Short Term Memory). Essa camada captura o histórico de interações dos usuários, proporcionando recomendações mais precisas e contextualizadas. Os resultados, avaliados quantitativamente via RMSE, Precisão@10 e Recall@10, evidenciam melhorias significativas em relação ao modelo base, demonstrando a eficácia da personalização para otimizar a experiência do visitante. O estudo contribui para o avanço do Smart Tourism, sugerindo caminhos futuros para integração de dados contextuais em tempo real e avaliação da experiência do usuário por meio de interfaces interativas.

**PALAVRAS-CHAVE:** machine learning; turismo inteligente; sistemas de recomendação; deep learning; tensorflow recommenders;

### INTRODUÇÃO

Segundo o IBGE, o índice de atividades turísticas no Brasil registrou um crescimento de 5,4% no primeiro trimestre de 2025 em comparação com o mesmo período do ano passado<sup>1</sup>. Estima-se que entre janeiro e março, mais de 23,7 milhões de passageiros circularam em voos domésticos nos aeroportos do país, e especialmente no Litoral Norte de São Paulo, um dos destinos mais procurados do estado, esse crescimento se traduz em um fluxo intenso de visitantes atraídos por sua vasta gama de atrativos que incluem praias, trilhas, gastronomia diversificada, artesanato local e eventos culturais. De acordo com Gretzel et al. (2015), o Smart Tourism, ou Turismo Inteligente, refere-se principalmente à integração da infraestrutura da cidade com as

<sup>1</sup>Disponível em: <https://www.gov.br/turismo/pt-br/assuntos/noticias/setor-de-turismo-mantem-alt-a-e-registra-crescimento-de-5-4-no-primeiro-trimestre-de-2025>

tecnologias que permitem que os turistas interajam com o local. Logo, é necessário que o recém chegado à cidade obtenha informações e descubra atividades de seu interesse com facilidade.

A pesquisa inicial demonstrou a viabilidade da filtragem de guias turísticos com base nas preferências dos usuários, sendo conduzida sob a perspectiva de que ferramentas digitais ampliam o acesso à informação turística e reduzem barreiras (AUTORES, 2025). O presente trabalho visa propor e validar um modelo de recomendação de duas etapas (recuperação e classificação) utilizando redes neurais profundas com a biblioteca TensorFlow Recommenders (TFRS), para atuar como assistente de criação de um roteiro turístico personalizado e individualizado.

## TEORIA

Para Ricci; Rokach; Shapira, (2022), sistemas de recomendação são algoritmos projetados para prever a “avaliação” ou “preferência” que um usuário daria a um item. O item pode ser um filme, um livro ou, no contexto deste trabalho, um ponto de interesse turístico. O objetivo de um sistema como este é filtrar informações e apresentar ao usuário os itens mais relevantes. Para construir um modelo de recomendação, modelos baseados em redes neurais profundas tornaram-se o estado da arte em estudos da área, pois elas são capazes de aprender padrões complexos e não-lineares a partir dos dados brutos. Uma técnica central nesses modelos é o uso de *embeddings*, que são vetores densos que representam usuários e itens. As principais abordagens para a filtragem dentro de um sistema recomendador são: filtragem baseada em conteúdo e filtragem colaborativa. “A filtragem baseada em conteúdo opera com base nas características dos itens e no perfil de preferência do usuário” (Peter Brusilovsky, 2007). A filtragem colaborativa, por outro lado, baseia-se na premissa de que usuários com comportamentos semelhantes no passado terão gostos semelhantes no futuro.

Diversos estudos recentes têm explorado o uso de aprendizado profundo em sistemas de recomendação turísticos, cada um adotando diferentes abordagens e técnicas. O trabalho de Shafqat; Byun (2020) utilizou uma rede LSTM para recomendações contextuais, considerando a sequência de locais visitados pelos usuários, mas sem explorar embeddings semânticos dos destinos. Semelhantemente, Cui; Liang; Zhang (2024) desenvolveu uma LSTM hierárquica em um cenário de contexto, também focando na ordem das visitas, porém sem trabalhar com representações semânticas complexas. Já Cepeda-Pacheco; Domingo (2022) proporam uma combinação de CNN e IoT para capturar dados contextuais em tempo real, embora sua dependência de sensores físicos torne a aplicação menos viável em ambientes sem infraestrutura adequada. Complementando, Flórez et al. (2025) integraram aprendizado profundo e ontologias, buscando enriquecer o contexto sem modelar a sequência temporal das interações. Diferentemente desses trabalhos, o presente estudo propõe uma abordagem híbrida baseada em embeddings e LSTM, incorporando tanto o histórico sequencial do usuário quanto uma personalização dinâmica, superando limitações observadas nas abordagens anteriores.

## MATERIAIS E MÉTODOS

A solução proposta para o estudo foi um modelo híbrido de recomendações que oferece ao turista sugestões de locais e atividades que correspondam genuinamente às suas preferências, histórico de interesses e comportamento. A abordagem se constitui em duas etapas, a recuperação e a classificação. A arquitetura do sistema foi pensada para capturar tanto as preferências iniciais como também a evolução dos interesses dos usuários, utilizando uma camada LSTM (Long Short Term Memory) desenvolvida inicialmente por Hochreiter; Schmidhuber (1997), para armazenar o histórico do turista. Dessa forma, o sistema é capaz de oferecer recomendações mais precisas e dinâmicas, contribuindo para reduzir a sobrecarga de opções e promover uma experiência turística mais personalizada. A Figura 1 ilustra algumas personas de usuários que se beneficiariam com o sistema.

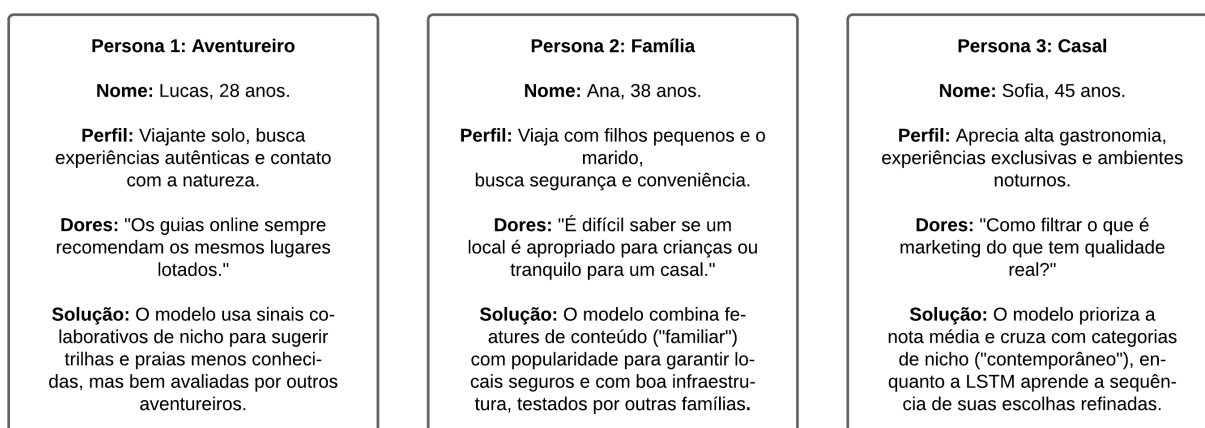


Figura 1: Personas de possíveis utilizadores do algoritmo, suas necessidades e a respectiva solução.

### Ferramentas e tecnologias utilizadas

A implementação de um sistema de recomendação moderno exige a análise das ferramentas disponíveis, cada uma contendo suas particularidades. Frameworks como o LightFM<sup>2</sup> são adequados para a prototipagem rápida de algoritmos clássicos de recomendação, como os baseados em fatoração de matrizes e k-NN. No entanto, eles apresentam limitações para a construção de arquiteturas de deep learning customizadas, que era um requisito deste estudo. No outro extremo, bibliotecas de baixo nível como o PyTorch<sup>3</sup> oferecem flexibilidade máxima para a criação de modelos complexos em larga escala, mas com um custo elevado em tempo de desenvolvimento e complexidade de implementação. Construída sobre o TensorFlow, a biblioteca TFRS abstrai grande parte da complexidade associada à criação de sistemas de recomendação TensorFlow 2025, o que a tornou a melhor escolha para este estudo.

<sup>2</sup>Disponível em: <https://github.com/lyst/lightfm>

<sup>3</sup>Disponível em: <https://pytorch.org>

## Dataset

Para o treinamento do modelo, foram utilizados dois datasets, um com dados reduzidos e utilizado em um ambiente controlado para validação da arquitetura e evoluindo para um cenário de maior complexidade e escala. Inicialmente, para a prototipagem, foi utilizado o dataset Dynamic Tourism Route Dataset (DTRD)<sup>4</sup>, que reúne informações de viajantes e detalhes de atrações. Este dataset foi fundamental para verificar a viabilidade da abordagem híbrida e realizar ajustes rápidos na arquitetura inicial. Contudo, sua escala limitada não seria suficiente para treinar um modelo generalizável, capaz de lidar com a esparsidade de dados encontrada em sistemas reais. O modelo final foi desenvolvido sobre o Yelp Dataset<sup>5</sup>, este fornece um sistema de avaliação baseado em uma escala de cinco estrelas, com usuários atribuindo notas e escrevendo avaliações com base em suas experiências pessoais. A migração para este dataset foi uma decisão deliberada para submeter o modelo a um cenário do mundo real, com uma vasta gama de usuários e itens, garantindo que os resultados fossem mais representativos e a solução final, mais escalável. Para o escopo deste trabalho, foi utilizado um subconjunto com mais de 500.000 avaliações.

## Desenvolvimento da solução

Durante a fase inicial da criação do algoritmo de recomendação, foram implementadas seis etapas para o pré-processamento dos dados. As etapas foram: (1) tratamento dos dados dos arquivos JSON, retirando linhas com informações vazias; (2) combinação dos arquivos utilizando o índice businessid; (3) seleção das colunas a serem utilizadas: userid, businessid, userrating, name, categories, reviewcount, city e avgbusinessrating; (4) conversão dos dados para o formato Dataset do Tensorflow; (5) divisão dos dados entre 80% de treino e 20% de teste; (6) definição dos vocabulários a serem utilizados pelo TFRS (useridvocab, businessidvocab, etc.). O processo completo da arquitetura do modelo pode ser visualizado na Figura 2.

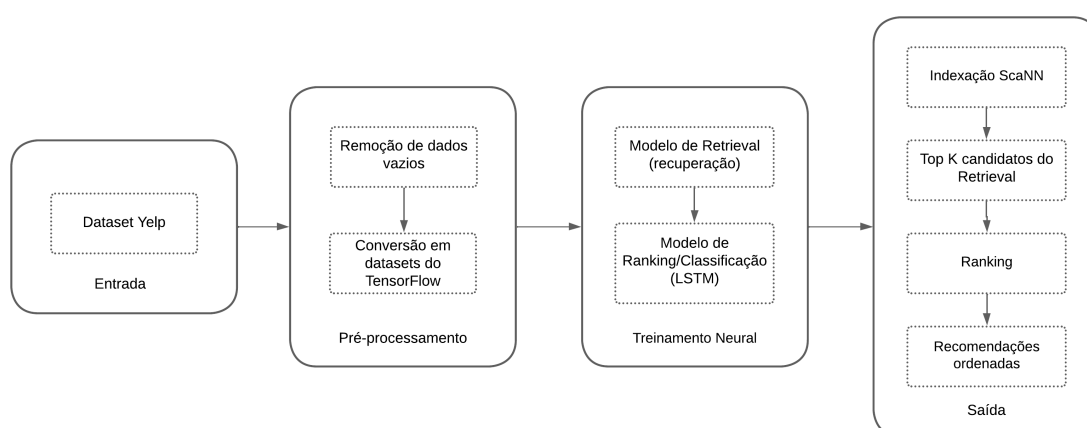


Figura 2: Diagrama do modelo de recomendação proposto.

A primeira etapa do treinamento, a recuperação, é responsável por selecionar um subcon-

<sup>4</sup>Disponível em: <https://www.kaggle.com/datasets/sudeepshouche/dynamic-tourism-route-datasetdtrd>

<sup>5</sup>Disponível em: <https://www.kaggle.com/datasets/yelp-dataset/yelp-dataset>

junto de candidatos relevantes. Para este fim, foi implementado um modelo de “duas torres”, que consiste em duas redes neurais independentes: a “torre do usuário” e a “torre do item”, que geram representações vetoriais (embeddings). O objetivo do treinamento é aprender a posicionar os vetores de itens com os quais um usuário interagiu mais próximos do vetor deste usuário. Essa lógica é implementada pela tarefa `tfrs.tasks.Retrieval`, que maximiza a similaridade entre os pares positivos (usuário-item que interagiram). O vetor de um dado usuário é usado para consultar eficientemente um índice de todos os vetores de itens, construído com `tfrs.layers.factorized_top_k.ScaNN`, que retorna rapidamente os candidatos mais promissores. A segunda etapa, de classificação, opera sobre os candidatos gerados e sua finalidade é prever uma pontuação de avaliação precisa para cada par usuário-item. Este modelo é treinado como um problema de regressão, incorporando features de conteúdo e contexto, como as categorias do negócio, a cidade e a contagem de avaliações. Sua implementação é gerenciada pela tarefa `tfrs.tasks.Ranking`.

Para capturar o histórico das preferências dos usuários, uma camada LSTM foi incorporada à arquitetura do modelo de classificação. No contexto deste estudo, a camada LSTM recebe como entrada a sequência cronológica dos embeddings dos negócios com os quais o usuário interagiu. Ao processar esta sequência, a rede aprende a modelar a evolução do gosto do usuário, e seu estado final serve como um vetor de representação dinâmico de todo o histórico. Por fim, este vetor é concatenado às demais features para a predição final da avaliação. Na fase de avaliação, o foco quantitativo foi no modelo de classificação, utilizando o conjunto de teste. A métrica principal selecionada foi o Erro Quadrático Médio da Raiz (RMSE), por mensurar diretamente a magnitude média dos erros de previsão, conforme estudado por Hyndman; Athanasopoulos (2018), seguindo a fórmula abaixo:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Portanto, um valor de RMSE reduzido servirá como o principal indicador do sucesso do modelo em prever com precisão as preferências dos usuários.

## RESULTADOS E DISCUSSÕES

O desenvolvimento do sistema seguiu uma abordagem iterativa, partindo de um modelo base no dataset de prototipagem (DTRD) para validar a arquitetura e progredindo para os modelos principais no dataset Yelp. A Tabela 1 apresenta os resultados do modelo inicial no ambiente controlado.

Tabela 1: Resultados do modelo de prototipagem no Dataset DTRD.

Métrica	Valor
RMSE	1.04
Precisão@10 <sup>6</sup>	0.65
Recall@10	0.49

Após a validação da arquitetura, o foco foi transferido para o dataset Yelp. Foi implementado um modelo base que combinava embeddings e features de conteúdo, mas sem considerar a sequência das interações. Os resultados deste modelo são mostrados na Tabela 2.

Tabela 2: Desempenho do modelo base no dataset Yelp (Sem LSTM).

Métrica	Valor
RMSE	0.95
Precisão@10	0.70
Recall@10	0.53

A principal inovação deste estudo foi a introdução da camada LSTM para processar o histórico de avaliações como uma sequência. O modelo final, treinado no mesmo dataset Yelp, incorporou essa camada para capturar a dinâmica dos interesses do usuário ao longo do tempo. Os resultados desta abordagem otimizada são apresentados na Tabela 3.

Tabela 3: Desempenho do modelo final no Dataset Yelp (com LSTM).

Métrica	Valor
RMSE	<b>0.87</b>
Precisão@10	<b>0.75</b>
Recall@10	<b>0.59</b>

A comparação direta entre a Tabela 2 e a Tabela 3 demonstra um avanço significativo. A redução do RMSE de 0.95 para 0.87 representa uma diminuição de 8,4% no erro médio de predição. A melhora mais expressiva, no entanto, é observada no ranking, com a Precisão@10 aumentando em 7,1% e o Recall@10 em 11,3%. Essa melhoria é atribuída diretamente à capacidade da camada LSTM de modelar a evolução dos interesses do turista, provando que a ordem das interações é uma informação valiosa que leva a recomendações mais relevantes.

## CONSIDERAÇÕES FINAIS

Este trabalho demonstrou com sucesso o desenvolvimento e a otimização de um modelo híbrido de recomendação para a personalização da experiência turística. Partindo de um modelo base validado em um ambiente controlado, a pesquisa evoluiu para uma solução robusta, utilizando deep learning e processamento sequencial com LSTM em um dataset de larga escala. O estudo contribui para a área de Turismo Inteligente, oferecendo uma arquitetura eficaz para mitigar o paradoxo da escolha e promover um turismo mais satisfatório. Para expandir o impacto e a sofisticação desta pesquisa, sugere-se para trabalhos futuros a integração de dados contextuais em tempo real, como condições climáticas, horário, dia da semana e a geolocalização do turista.

Em síntese, o sistema proposto já demonstra seu potencial como solução inovadora, unindo um problema real com técnicas avançadas de machine learning. O presente estudo utilizou apenas dados públicos e anônimos, não exigindo que o estudo fosse aprovado pelo Comitê de Ética em Pesquisa (CEP-IFSP). Para pesquisas futuras envolvendo informações

peçoais ou dados sensíveis de turistas, poderá ser necessária a observância da Lei Geral de Proteção de Dados Pessoais (LGPD) e outras normas institucionais.

**Agradecimento:** Os autores agradecem ao Centro de Pesquisas CEPIN-CIS pela viabilização do estudo, bem como pelo suporte técnico e disponibilização de recursos essenciais para a realização deste trabalho. A todos que participaram direta ou indiretamente deste trabalho.

## REFERÊNCIAS

- AUTORES. Integrando dados abertos pertinentes ao turismo à assistentes virtuais. **CONICT**, 2025.
- CEPEDA-PACHECO, J. C.; DOMINGO, M. C. Deep learning and Internet of Things for tourist attraction recommendations in smart cities. **Neural Computing and Applications**, 2022.
- CUI, Y.; LIANG, S.; ZHANG, Y. Multimodal representation learning for tourism recommendation with two-tower architecture. **PLOS ONE**, v. 19, n. 2, 2024. DOI: <https://doi.org/10.1371/journal.pone.0299370>.
- FLÓREZ, M. et al. A Context-Aware Tourism Recommender System Using a Hybrid Method Combining Deep Learning and Ontology-Based Knowledge. **Journal of Theoretical and Applied Electronic Commerce Research**, v. 20, n. 3, 2025. DOI: 10.3390/jtaer20030194. Disponível em: <https://www.mdpi.com/0718-1876/20/3/194>.
- GRETZEL, U. et al. Smart Tourism: Foundations and Developments. **Electronic Markets**, v. 25, n. 3, p. 179–188, 2015.
- HOCHREITER, S.; SCHMIDHUBER, J. Long Short-Term Memory. **Neural Computation**, MIT Press, v. 9, n. 8, p. 1735–1780, 1997.
- HYNDMAN, R. J.; ATHANASOPOULOS, G. **Forecasting: Principles and Practice**. [S.l.]: OTexts, 2018. Disponível em: <https://otexts.com/fpp3/>.
- PETER BRUSILOVSKY ALFRED KOBASA, W. N. **The Adaptive Web: Methods and Strategies of Web Personalization**. [S.l.]: Springer, 2007.
- RICCI, F.; ROKACH, L.; SHAPIRA, B. **Recommender Systems Handbook**. [S.l.]: Springer US, 2022. (Computer Science). ISBN 9781071621974. Disponível em: <https://books.google.com.br/books?id=dWBsEAAAQBAJ>.
- SHAFQAT, W.; BYUN, Y.-C. A Context-Aware Location Recommendation System for Tourists Using Hierarchical LSTM Model. **Sustainability**, v. 12, n. 10, 2020. DOI: 10.3390/su12104107. Disponível em: <https://www.mdpi.com/2071-1050/12/10/4107>.
- TENSORFLOW. **TensorFlow Recommenders (TFRS) Overview**. [S.l.: s.n.], 2025. Acesso em: 12 out. 2025. Disponível em: <https://www.tensorflow.org/recommenders>.