**Abstract:**

Abstract: Fashion is an ever-evolving field where inspiration and personalization are key factors for user interaction. However, nowadays existing digital tools do not offer dynamic and personalized recommendations that combine visual understanding and interaction with natural language. To address this shortcoming, I'm introducing a novel AI-based system that leverages both deep learning techniques in computer vision and natural language processing (NLP) to improve the fashion recommendation experience. Users who access it can interact with the chatbot through conversational queries and images to discover similar outfits and garments that fit their preferences, achieving greater combinations of that garment with other different ones. In addition, the system estimates the similarity of garments based on visual and semantic characteristics providing direct links to e-commerce platforms to immediately access the recommended products and to be able to purchase them knowing in advance their characteristics (price, color, description, etc.). This approach not only improves the relevance and intuitiveness of fashion recommendations but also reduces the lack of systems that include user intent and product discovery.

Index Terms—Fashion Recommendation, Natural Language Processing, Computer Vision, Chatbots, Deep Learning, Visual Similarity, Personalized Recommendation Systems, E-commerce

**Introducción:**

La moda es un fenómeno cultural y social que evoluciona constantemente, y que está influenciado por tendencias, contextos históricos y preferencias personales. Actualmente, con la expansión de las redes sociales y el comercio electrónico, las personas buscan cada vez más inspiración y asesoramiento para definir su estilo y seleccionar sus prendas de vestir, dependiendo de la ocasión a la que deban acudir. Sin embargo, la mayoría de las plataformas existentes carecen de sistemas personalizados, visualmente interactivos y fáciles de entender para asistir al usuario de forma efectiva en este proceso de sugerencia y elección. Es por este motivo que muchas personas deciden dejar a un lado la parte tecnológica y contratar a personas especializadas en el sector de la moda, pero lo que no saben es que sin la tecnología ellas no podrían desarrollar correctamente su trabajo.

El proceso de elección de un conjunto de ropa involucra múltiples factores, como son la estética personal, la ocasión de uso, la combinación de colores, las tendencias actuales y la disponibilidad de productos en el mercado. Sin embargo, las soluciones existentes, basadas principalmente en motores de búsqueda convencionales o filtros categóricos, presentan limitaciones importantes, ya que carecen de herramientas capaces de interpretar tanto el lenguaje natural como las imágenes de todo tipo de prendas.

Hoy en día existen catálogos online y motores de búsqueda visual, pero estos suelen ser limitados en cuanto a personalización, comprensión semántica y capacidad de sugerir alternativas relevantes. Por ello, si no disponen de trabajadores presenciales que ayuden a cada usuario, el trabajo que venden no es completamente asesorado y tampoco aprovechan plenamente las capacidades de análisis visual que proporciona la inteligencia artificial moderna.

Uno de los desafíos más relevantes en la recomendación de moda radica en el denominado "gap semántico"[1]: la discrepancia entre la descripción textual de una prenda y su representación visual. Superar este problema es fundamental para ofrecer recomendaciones realmente precisas y satisfactorias, basándose en los gustos de cada usuario y en todos los factores que radican en ese outfit.

En este contexto, el trabajo que he realizado propone el desarrollo de un chatbot inteligente que integra procesamiento de lenguaje natural (NLP), análisis de imágenes para recomendar outfits y prendas similares de manera automatizada (Computer Vision and Advanced Machine Learning). Además de sugerir combinaciones personalizadas, el sistema calcula la similitud entre distintas prendas basándose en detalles visuales y semánticos, y proporciona enlaces directos hacia sitios web donde los usuarios pueden adquirir los productos sugeridos sabiendo previamente todas sus características, entre ellas su precio (Intelligent Systems).

Las principales contribuciones que se han desarrollado para este proyecto son:

* La integración de técnicas de NLP como procesamiento de textos y uso de transformers, para la completa comprensión del usuario.
* La interpretación visual del completo dataset junto con sus características clave para la recomendación de todo tipo de prendas.
* La creación de un sistema de búsqueda de prendas similares a partir de imágenes y descripciones basado en un grafo de similaridad.
* El diseño de una experiencia interactiva que conecta la conversación con un chatbot, las recomendaciones personalizadas y la vinculación con plataformas de compra online para poder adquirir dichas prendas en su web.

1. <https://www.tdx.cat/bitstream/handle/10803/369847/Tesi_Jordi_Iglesias.pdf?sequence=>

**Estado del arte:**

**A. Sistemas de Recomendación en Moda**

La mezcla de la inteligencia artificial con la moda ha crecido significativamente como un área de investigación y esto ha conseguido mejoras en el campo del comercio. La integración de visión por computador y procesamiento del lenguaje ha sido clave para el trabajo de sistemas de recomendación basados principalmente en análisis de imágenes o en sugerencias basadas en historial de usuario.

En el área de la visión por computador, técnicas como las redes convolucionales (CNNs) han sido ampliamente utilizadas para extraer características visuales de prendas y calcular similitudes entre ellas. Por otra parte, modelos como FashionNet y DeepFashion[2] han permitido avances significativos en el reconocimiento y la clasificación de ropa a partir de imágenes, estableciendo bases de datos estandarizadas que han impulsado el desarrollo de nuevos algoritmos. Aun así, los datasets existentes no contienen las características completas y necesarias para trabajar con toda la información en este campo.

Paralelamente, el procesamiento de lenguaje natural ha sido explorado para mejorar la interacción entre los usuarios y los sistemas de recomendación, con el fin de conseguir una propuesta completa. Estudios recientes han empleado modelos basados en transformers, como BERT, para interpretar descripciones de prendas y consultas de usuarios, mejorando así la relevancia de las sugerencias [3]. Hoy en día, estos modelos son la mejor opción a la hora de trabajar con descripciones, permitiendo el procesamiento de la frase palabra por palabra, y en consecuencia, solucionando el problema del tiempo de procesado.

1. <https://paperswithcode.com/dataset/deepfashion>
2. <https://www.teldat.com/es/blog/ai-inteligencia-artificial-bert/>

**B. Recomendación Multimodal y Chatbots**

Los sistemas de recomendación multimodales aprovechan e integran múltiples tipos de datos para predecir y sugerir elementos que se ajusten a sus preferencias [4]. De hecho, gracias a combinar información textual y visual, han mostrado resultados prometedores. Sin embargo, gran parte de estos sistemas están diseñados para contextos de búsqueda pasiva, donde el usuario ingresa palabras clave o selecciona filtros predefinidos, en lugar de interactuar de manera conversacional. Simplemente son usados para buscar en un gran contexto palabras clave con el fin de recopilar información, pero no con el de guardarla.

Adicionalmente, el uso de chatbots en el ámbito comercial ha ganado popularidad, especialmente en el sector de ventas online. No obstante, la mayoría de los chatbots actuales se limitan a funciones básicas de atención al cliente y carecen de capacidades avanzadas de interpretación de imágenes o de recomendaciones personalizadas basadas en el contenido visual. De hecho, la mayoría de estos no tiene la capacidad de entender frases completas y por ello esos sistemas muestran diferentes opciones para que el usuario seleccione una y se envíe, sin la necesidad de escribir ningún tipo de oración.

1. <https://paperswithcode-com.translate.goog/task/multimodal-recommendation?_x_tr_sl=en&_x_tr_tl=es&_x_tr_hl=es&_x_tr_pto=rq>

**C. Brechas y Oportunidades**

A pesar de estos avances, sigue existiendo una gran brecha en el desarrollo de sistemas capaces de comprender lenguaje natural de forma conversacional, analizar imágenes de moda de manera precisa y proporcionar recomendaciones personalizadas que integren ambos tipos de información. Esta falta motiva el desarrollo de este proyecto, que busca combinar técnicas avanzadas de procesamiento de lenguaje natural y visión por computadora en un chatbot inteligente que mejore la experiencia de investigación y compra de moda para los usuarios.

**Experimentación:**

***\subsection{A. Creación del Dataset}***

Para este proyecto he desarrollado un dataset propio, debido a la inexistencia de un conjunto de imágenes de ropa claro y con todas las características necesarias para trabajar con él sin que presente problemas. El conjunto de datos creado es específico y adaptado a los objetivos de recomendación personalizada de moda. De hecho, está compuesto por aproximadamente 300 y 500 imágenes de prendas, recopiladas manualmente desde la plataforma web de Zara, combinando procesos manuales y técnicas de web scraping para extraer tanto los archivos visuales como las descripciones de los productos.

Las imágenes se han organizado en once categorías principales: tacones, accesorios, botas, camisetas, camisas, sudaderas, jerséis, pantalones, zapatillas, chaquetas y abrigos. Cada imagen fue almacenada en formato PNG para preservar la calidad visual y facilitar posteriores etapas de procesamiento de imagen. Adicionalmente, se crearon archivos CSV asociados, en los cuales se registraron las etiquetas correspondientes y las descripciones textuales extraídas, permitiendo así la vinculación entre los datos visuales y semánticos.

***\subsection{B. Procesamiento de Datos}***

El procesamiento de datos realizado para este proyecto abarcó tanto los elementos visuales como los textuales.

Para el procesamiento de imágenes, se aplicaron procesos de lectura, conversión de color RGB, normalización de píxeles y redimensionamiento a un tamaño estándar de píxeles. Estas imágenes fueron utilizadas para el entrenamiento de dos redes convolucionales de transfer learning: VGG16[5] y ResNet50[6], ambas preentrenadas sobre ImageNet. La salida de la capa de características de VGG16 fue empleada para generar representaciones vectoriales de las imágenes, que sirvieron para realizar clustering mediante KMeans y para la construcción de un grafo de similitud, donde se conectaban las prendas más similares.

Respecto a los datos textuales, se obtuvieron de revistas de moda como Vogue o Elle, y de artículos online para los cuales fue necesario aplicar web scrapping. Se utilizó la librería spaCy para realizar tokenización, eliminación de stopwords, lematización y normalización de las descripciones extraídas. A partir de estos textos preprocesados, se aplicó TF-IDF para la extracción de palabras clave relevantes en el dominio de la moda, y se usaron modelos basados en BERT multilingüe para analizar tendencias semánticas dentro del corpus textual. Debido a la baja existencia de revistas de moda con renombre en inglés, fue necesario descargar diferentes ejemplares y traducirlos para que fuera posible el análisis de las palabras.

Adicionalmente, se implementó un sistema de generación automática de descripciones de outfits basado en modelos generativos como GPT-2, tanto a partir de imágenes procesadas como de preferencias explícitas proporcionadas por los usuarios.

Por último, para construir un sistema de recomendación basado en retroalimentación, se generó un conjunto sintético de valoraciones de usuario sobre artículos de moda y se entrenó un modelo de factorización de matrices mediante la técnica SVD (Singular Value Decomposition)[7], aprobado con una metodología de validación cruzada.

Todo el conjunto de datos procesados y sus respectivas características fueron almacenados en formato CSV y NPY para su posterior integración en el motor de recomendación conversacional.

1. <https://keepcoding.io/blog/arquitectura-vgg16-vgg19-deep-learning/>
2. <https://blog.roboflow.com/what-is-resnet-50/>
3. <https://www.geeksforgeeks.org/singular-value-decomposition-svd/>

**\subsection{C. Arquitectura del sistema}**

**\subsection{A. Chatbot Inteligente Basado en Rasa}**

Diagrama

El contenido generado por IA puede ser incorrecto.

La arquitectura general del sistema se diseñó con el objetivo de ofrecer una experiencia de recomendación de moda personalizada a través de una interacción conversacional fluida y fácil de entender. El sistema integra procesamiento de lenguaje natural, visión por computador y mecanismos de recomendación en un único flujo operativo, aplicando tanto sistemas inteligentes como aprendizaje automático avanzado.

Para la implementación del componente conversacional, se utilizó la plataforma Rasa, que es un framework de código abierto y de inteligencia artificial asincrónica empleado principalmente para la creación de asistentes contextuales, es decir, aplicaciones de bot de chat [8]. Este framework fue entrenado para interpretar intenciones de usuario, reconocer entidades relevantes y gestionar diálogos dinámicos.

Rasa se compone principalmente de dos módulos fundamentales: Rasa NLU (Natural Language Understanding) y Rasa Core [9].

* \textbf{Rasa NLU} se encarga de procesar las entradas textuales de los usuarios, identificando las \textit{intenciones} y extrayendo \textit{entidades} relevantes. Para este proyecto, el entrenamiento de NLU se apoyó en un modelo basado en BERT, optimizando la comprensión semántica de las consultas de moda en lenguaje natural.
* \textbf{Rasa Core} gestiona la conversación y la lógica del diálogo a través de un conjunto de \textit{historias} y \textit{reglas}. Las historias representan flujos conversacionales de ejemplo que el asistente debe aprender a manejar, mientras que las reglas definen respuestas específicas ante condiciones particulares.

La configuración del chatbot se estructuró en los siguientes componentes:

\begin{itemize}

\item \textbf{Domain}: Define los intents, entidades, acciones y respuestas disponibles en el asistente.

\item \textbf{NLU}: Contiene los ejemplos de entrenamiento para la detección de intents y extracción de entidades.

\item \textbf{Stories}: Describe los posibles caminos conversacionales basados en secuencias de intents y acciones.

\item \textbf{Rules}: Establece acciones determinísticas ante situaciones específicas, como la confirmación de selección de prenda.

\item \textbf{Actions}: Define acciones personalizadas, como la generación de descripciones de outfits.

\end{itemize}

Durante la interacción, cuando el usuario envía un mensaje de texto, este se procesa a través del pipeline de NLU, identificando la intención y las entidades. El sistema, a partir de esa comprensión, proporciona recomendaciones de outfits, sugerencias de estilo y permite acceder a los dos otros apartados de recomendación por imágenes y grafos de similitud para tener información sobre prendas específicas.

1. <https://techdocs.broadcom.com/es/es/ca-enterprise-software/it-operations-management/dx-apm-saas/SaaS/python-agent/Python-Agent-Extensions/Rasa.html>
2. <https://legacy-docs-oss.rasa.com/docs/rasa/>

**\subsection{B. Recomendación visual de prendas}**

Además de la interacción conversacional, el sistema incluye un módulo de recomendación basado en la subida de imágenes de prendas. Cuando el usuario proporciona una imagen, esta es procesada mediante una red convolucional ResNet50, preentrenada sobre ImageNet, con el fin de extraer características visuales profundas. Se ha escogido esta red convolucional debido a que permite entrenar redes muy profundas (de más de 100 capas), controlando con éxito el problema de ‘desvanecimiento *de gradiente’* [10].

Estas representaciones visuales son comparadas utilizando un grafo de similitud previamente construido, permitiendo encontrar artículos similares a la prenda subida por el usuario. El grafo se construyó a partir de características extraídas de un conjunto de prendas organizadas mediante técnicas de clustering (KMeans), optimizando así la navegación por relaciones visuales. Además, el sistema ofrece una generación automática de descripciones de outfits basada en modelos de lenguaje generativos como GPT-2, adaptando las sugerencias a las preferencias de estilo, color y ocasión declaradas.

El motor de recomendación multimodal se complementa mediante un modelo de factorización de matrices SVD, que permite sugerir prendas en función de valoraciones simuladas de usuarios, incorporando así un enfoque colaborativo que enriquece las recomendaciones. Este modelo se basa en una generalización de la descomposición en autovalores de una matriz, y se puede aplicar a cualquier matriz rectangular, no solo a las matrices cuadradas. Sirve principalmente para encontrar patrones subyacentes en grandes textos [5].

Esta integración de componentes permite al sistema ofrecer una experiencia personalizada basada en entradas visuales, para finalmente proponer outfits completos, prendas similares o estilos alternativos, mejorando la experiencia de descubrimiento de moda de manera integral.

1. <https://interactivechaos.com/es/wiki/descomposicion-en-valores-singulares-svd>

**\subsection{C. Grafo de similitud de prendas}**

Para mejorar la calidad de las recomendaciones visuales y permitir una exploración interactiva de prendas similares, se diseñó y construyó un grafo de similitud basado en las características visuales extraídas de las imágenes de prendas.

El proceso de construcción del grafo se inició extrayendo embeddings visuales utilizando una red convolucional ResNet50 preentrenada, aplicada a las imágenes normalizadas del dataset. A partir de estos embeddings, se calcularon las similitudes entre pares de prendas utilizando la métrica de cosine similarity. Esta métrica está basada en la medición del ángulo entre dos documentos en el espacio métrico de múltiples dimensiones [11].

El grafo de similitud se construyó siguiendo los siguientes pasos:

\begin{itemize}

\item Cada nodo representa una prenda del dataset, etiquetada con su categoría.

\item Para cada nodo, se conectaron las \(k\) prendas más similares, donde \(k\) es un parámetro configurable, que establece las conexiones mediante aristas a lo largo del grado de similitud.

\end{itemize}

La construcción del grafo se realizó mediante la librería NetworkX, y su almacenamiento persistente se implementó utilizando la serialización mediante Pickle, permitiendo su posterior carga y utilización sin necesidad de recalcular las similitudes. El principal uso de este es convertir un objeto Python en un flujo de bytes, que puede almacenarse en un archivo o transferirse a través de una red [12].

Para recomendar prendas similares a partir de una prenda de entrada, se implementó un algoritmo de búsqueda en anchura Breadth-First Search (BFS) sobre el grafo. Este algoritmo permite explorar nodos vecinos hasta una profundidad determinada, recuperando así un conjunto de prendas altamente relacionadas con la prenda inicial del grafo.

Además, se desarrolló una visualización interactiva del grafo utilizando Plotly, permitiendo al usuario explorar subgrafos locales centrados en una prenda específica. De esta manera el usuario puede saber visualmente. La visualización incluye:

\begin{itemize}

\item Representación de nodos (prendas) y aristas (similitudes).

\item Relación directa entre las diferentes prendas.

\item Colores diferenciados por cada categoría de prenda.

\item Color destacado al buscar una determinada prenda.

\item Imágenes y características mostradas dinámicamente al seleccionar un nodo concreto.

\end{itemize}

Este componente proporciona una herramienta poderosa para la navegación visual dentro del apartado de las prendas, permitiendo al usuario descubrir combinaciones basadas en relaciones de similitud profunda y obteniendo más información sobre cada uno de los diferentes ítems.

1. <https://www.datahack.es/motores-de-recomendacion-con-python-parte-2/>
2. <https://docs.python.org/es/3.13/library/pickle.html>

**\subsection{D. Configuración experimental}**

Para llevar a cabo la implementación del sistema propuesto, se utilizaron diferentes frameworks y herramientas de aprendizaje automático y procesamiento de lenguaje natural. El entorno de desarrollo principal estuvo basado en Python 3.10 debido a la conexión de esta versión con Rasa, empleando bibliotecas como TensorFlow, Keras, scikit-learn, Hugging Face Transformers, NetworkX, Plotly, Rasa y Streamlit.

En cuanto al procesamiento de imágenes, se entrenó un modelo de clasificación de prendas utilizando la arquitectura VGG16 con pesos preentrenados en ImageNet. El modelo se entrenó congelando la base convolucional y añadiendo capas densas personalizadas para la clasificación de 11 categorías de prendas.

Para la detección de estilo, se utilizó un segundo modelo basado en ResNet50, también preentrenado y adaptado con una capa densa final para clasificar 5 estilos (casual, formal, sportive, elegant, urban).

El sistema de recomendación fue implementado mediante la librería \textit{Surprise}, utilizando la técnica de \textit{Singular Value Decomposition} (SVD). Se generaron valoraciones sintéticas entre usuarios y prendas, y se aplicó validación cruzada (\textit{cross-validation}) con 5 particiones para evaluar el rendimiento general.

El componente de procesamiento de lenguaje natural fue gestionado mediante la plataforma Rasa, combinada con modelos de BERT multilingüe para la detección de intenciones. La generación de texto se llevó a cabo utilizando el modelo GPT-2 a través de la librería Transformers de Hugging Face.

Las métricas utilizadas para evaluar el sistema incluyen:

\begin{itemize}

\item \textbf{Accuracy} del modelo de clasificación de prendas.

\item \textbf{Precision} en la predicción de estilo.

\item \textbf{Top-K similarity} para las recomendaciones visuales.

\item \textbf{RMSE} en la validación cruzada del sistema colaborativo.

\item \textbf{Interacción y comprensión conversacional} en pruebas con usuarios.

\end{itemize}

Todos los modelos y pruebas se llevaron a cabo en un entorno de desarrollo local, utilizando la plataforma Rasa como núcleo del sistema conversacional.

**\section{ANÁLISIS DE RESULTADOS}**

Durante la fase experimental, se realizaron distintas pruebas para evaluar el rendimiento de cada módulo del sistema propuesto, abarcando desde la clasificación de prendas hasta la recomendación personalizada basada en imagen y texto.

En primer lugar, con el modelo de clasificación de prendas implementado con VGG16 se logró una precisión del 99.08\%, lo cual indica un excelente resultado en la tarea de identificar el tipo de prenda a partir de imágenes. Esto confirma que, pese al tamaño reducido del dataset, el modelo fue capaz de aprender representaciones visuales para las 11 categorías.

Por otra parte, el modelo de clasificación de estilo, basado en una red ResNet50 ajustada, no alcanzó tan buenos resultados tras 30 épocas de entrenamiento. La precisión obtenida fue del 5.81\%, lo que muestra una alta complejidad en la tarea de clasificación de estilos que se plantea como una línea de mejora para futuras actualizaciones del programa.

En cuanto a la organización visual, se extrajeron las características de las imágenes procesadas para construir un grafo de similitud, sobre el cual se aplicó clustering no supervisado, en concreto KMeans, en el que se obtuvieron cinco agrupaciones principales. Este grafo permitió realizar recomendaciones basadas en similitud visual y facilitó la exploración del espacio de prendas a través de algoritmos de búsqueda como BFS.

En cuanto al sistema de recomendación colaborativa, se implementó utilizando el algoritmo SVD sobre datos sintéticos de valoraciones. El modelo obtuvo un \textit{Root Mean Square Error} (RMSE) promedio de 1.4157 y un \textit{Mean Absolute Error} (MAE) de 1.2170, lo que representa un rendimiento bueno considerando las valoraciones y el tamaño del conjunto de usuarios que mostraron sus opiniones.

Finalmente, aunque no se presentan métricas formales del componente conversacional ni de la generación de descripciones con GPT-2, ambos fueron evaluados de manera cualitativa. Durante las pruebas, el chatbot demostró capacidad para interpretar correctamente intenciones en la mayoría de los casos y generar respuestas coherentes. Las descripciones de outfits generadas automáticamente también mostraron ser comprensibles y útiles para acompañar las recomendaciones.

En conjunto, todos estos resultados muestran un sistema funcional y bien estructurado con resultados prometedores, que cuenta con un pequeño margen de mejora en cuanto a la clasificación de estilos y a la generalización de las recomendaciones colaborativas.

**\section{CONCLUSIONES Y TRABAJOS FUTUROS}**

El desarrollo de este proyecto ha demostrado la viabilidad de integrar diferentes técnicas de inteligencia artificial, como son el procesamiento de lenguaje natural, visión por computador y sistemas de recomendación, en una plataforma unificada de asistencia conversacional para el mundo de la moda. La arquitectura propuesta permite a los usuarios interactuar de forma natural, tanto mediante texto como a través de imágenes, recibiendo recomendaciones personalizadas y visualmente coherentes.

El sistema logró una alta precisión en la clasificación de prendas utilizando VGG16, así como una estructura de recomendación visual basada en grafos que facilita la exploración entre productos similares. También se incorporó un motor de recomendación colaborativa que, aunque basado en valoraciones sintéticas, mostró resultados consistentes. A nivel conversacional, la integración de Rasa permitió mantener diálogos fluidos, mientras que la generación de descripciones automáticas añadió un valor comunicativo al sistema. Todo en su conjunto mostró una combinación perfecta de técnicas de comunicación y recomendación destacable en un sector como es el de la moda.

No obstante, es cierto que existen algunos aspectos en los que se pueden realizar ciertas mejoras. El modelo encargado de clasificar estilos mostró un bajo rendimiento, aunque en la plataforma se han concluido resultados favorables, lo que señala la necesidad de ampliar el dataset y trabajar con ejemplos. Además, sería interesante evaluar el sistema con usuarios reales, recogiendo su feedback para que permita medir sus opiniones reales en un entorno práctico.

Como trabajos futuros, he planteado optimizar el módulo de estilo utilizando técnicas de fine-tuning más específicas, enriquecer el dataset con mayor cantidad de imágenes y añadir descripciones reales de mayor diversidad. También he considerado el despliegue del sistema en una plataforma web en producción, permitiendo su uso abierto y la recogida de datos para retroalimentar el aprendizaje.

En definitiva, este proyecto se muestra como una base sólida para la creación de asistentes virtuales aplicados al sector moda, combinando tecnologías emergentes con una interfaz accesible, funcional y sobre todo visual.