



Universidad
Nacional de
General
Sarmiento

ABRIL 2025

Laboratorio de Construcción de Software

TP Inicial - Parte final

Prototipo completo y documentación final:
Recomendación de grupos basada en habilidades

PRESENTADO POR:

GRUPO 3

Alfaro, Ezequiel

Ibacache, Maria Jose

Torales, Cecilia

Silva, Agustina

DOCENTES

Dikenstein, Leandro

Orozco, Francisco



Documentación Técnica: Modelo de Agrupamiento y Recomendación de Cursos

Introducción

Este informe presenta un modelo basado en el algoritmo de clustering K-Means, diseñado para agrupar empleados según dos características clave: su habilidad técnica y su nivel de experiencia. A partir de esto, se asigna una recomendación de curso específica para cada grupo, permitiendo una asignación de recursos educativos más efectiva.

El modelo utiliza técnicas de preprocesamiento de datos, como la codificación manual de variables categóricas y el escalado de características, para garantizar que las entradas sean compatibles con el algoritmo k-Means. Además, se incluye una visualización gráfica que facilita la interpretación de los resultados y la identificación de patrones en los datos.

Este documento detalla el funcionamiento del modelo, proporciona instrucciones claras para su implementación y describe cómo realizar predicciones para nuevos empleados.

Descripción del Modelo

El modelo implementado en el código proporcionado utiliza un enfoque de aprendizaje no supervisado basado en k-Means. El objetivo es agrupar empleados según dos características principales: su habilidad técnica (codificada manualmente) y su nivel de experiencia. Una vez que los empleados son agrupados, se asigna una recomendación de curso específica para cada cluster.

Componentes del Modelo

1. Codificación de Habilidades: Las habilidades técnicas ("Assembler", "Java", "Python", "SQL") son codificadas manualmente mediante un mapeo numérico.
2. Escalado de Datos: Las columnas Habilidad_Encoded y Nivel_Encoded son escaladas utilizando MinMaxScaler para normalizar los valores entre 0 y 1. Esto asegura que las características tengan la misma escala durante el proceso de clustering.
3. Clustering con k-Means: Se utiliza el algoritmo k-Means para agrupar a los empleados en `n_clusters` clusters. El número de clusters está definido como 12 en este caso.
4. Recomendaciones de Cursos: Cada cluster se asocia con un curso específico, lo que permite recomendar cursos personalizados para los empleados en función de su habilidad técnica y nivel de experiencia.
5. Visualización: Los resultados del clustering son visualizados en un gráfico de dispersión, donde cada punto representa a un empleado. Los centroides de los clusters también se grafican para mayor claridad.

Documentación Técnica: Modelo de Agrupamiento y Recomendación de Cursos

Instrucciones para Ejecutar el Código

Requisitos Previos:

1. Entorno Python: Asegúrese de tener instalado Python (versión 3.6 o superior).
2. Bibliotecas Necesarias: Instale las bibliotecas requeridas ejecutando el siguiente comando: `pip install pandas matplotlib scikit-learn numpy`
3. Archivo de Datos: El archivo `data.csv` debe estar presente en el directorio de trabajo. Este archivo debe contener al menos las siguientes columnas: *Habilidad*, con la habilidad técnica del empleado (valores posibles: "Assembler", "Java", "Python", "SQL") y *Nivel*, con el nivel de experiencia del empleado (valor numérico)

Pasos para Ejecutar el Código

- a. Cargar el Código: Guarde el código proporcionado en un archivo Python, por ejemplo, `modelo_kmeans.py`.
- b. Colocar el Archivo de Datos : Asegúrese de que el archivo `data.csv` esté en el mismo directorio que el script Python.
- c. Ejecutar el Script: Ejecute el script desde su IDE de preferencia o línea de comandos
- d. Resultados: Al finalizar la ejecución, se generará un gráfico de dispersión que muestra los clusters y sus respectivas recomendaciones de cursos.

Proceso de Entrenamiento del Modelo

Preprocesamiento de Datos

1. Codificación Manual de Habilidades: Las habilidades técnicas son mapeadas a valores numéricos utilizando el diccionario `habilidad_map`. Esta codificación se aplica a la columna `Habilidad` del DataFrame empleados.
2. Escalado de Características: Las columnas `Habilidad_Encoded` y `Nivel_Encoded` son escaladas utilizando `MinMaxScaler` para normalizar los valores entre 0 y 1.

Entrenamiento del Modelo

1. Selección de Características: Se seleccionan las columnas `Habilidad_Encoded` y `Nivel_Encoded` como características de entrada para el modelo.
2. Configuración del Modelo: Se define el número de clusters (`n_clusters`) y se inicializa el modelo k-Means
3. Entrenamiento: El modelo se entrena utilizando el método `fit_predict`, que asigna un cluster a cada empleado

Documentación Técnica: Modelo de Agrupamiento y Recomendación de Cursos

Proceso de Predicción

Una vez entrenado el modelo, se pueden realizar predicciones para nuevos empleados siguiendo estos pasos:

1. Preprocesamiento de Nuevos Datos:
 - Codifique las habilidades técnicas de los nuevos empleados utilizando el mismo mapeo (habilidad_map).
 - Escale las características Habilidad_Encoded y Nivel_Encoded utilizando la misma función MinMaxScaler utilizado durante el entrenamiento.
2. Predicción del Cluster:
 - Utilice el método predict del modelo k-Means para asignar un cluster a los nuevos empleados
3. Recomendación de Curso :
 - Asigne una recomendación de curso al nuevo empleado basándose en el cluster predicho

Visualización de Resultados

El código incluye una visualización gráfica de los clusters utilizando *matplotlib*. Cada cluster está representado por un color único, y los centroides de los clusters se destacan con un símbolo especial. Además, se añade una leyenda que muestra la recomendación de curso asociada a cada cluster.

Personalización de la Visualización

- Colores Dinámicos: Los colores de los clusters se generan dinámicamente utilizando el mapa de colores tab10.
- Jitter: Se aplica un pequeño jitter (ruido aleatorio) a los puntos para evitar superposiciones en la visualización.
- Leyenda Externa: La leyenda se coloca fuera del gráfico para mejorar la legibilidad.

Consideraciones Finales

1. Número de Clusters: El número de clusters está configurado como 12. Este valor puede ajustarse según el análisis exploratorio de datos (por ejemplo, utilizando el método del codo) para un resultado más óptimo.
2. Interpretación de Resultados: Las recomendaciones de cursos están predefinidas y asignadas manualmente a cada cluster. Si se cambia el número de clusters, será necesario actualizar el diccionario etiquetas_clusters.
3. Escalabilidad: El modelo es escalable para conjuntos de datos más grandes, pero el tiempo de entrenamiento puede aumentar con el tamaño del dataset.

Documentación Técnica: Modelo de Agrupamiento y Recomendación de Cursos

Conclusión

El desarrollo e implementación del modelo de clustering basado en k-Means presentado en este informe demuestra el potencial de las técnicas de aprendizaje no supervisado para abordar desafíos prácticos en la gestión de talento y la formación continua. Mediante la agrupación de empleados según sus habilidades técnicas y niveles de experiencia, el modelo permite ofrecer recomendaciones de cursos personalizadas que se alinean con las necesidades individuales y colectivas de los trabajadores. Este enfoque no solo optimiza la asignación de recursos educativos, sino que también contribuye a mejorar la satisfacción laboral, la productividad y el desarrollo profesional dentro de una organización.

Uno de los aspectos más destacados del modelo es su capacidad para procesar datos de manera eficiente y generar resultados interpretables. La codificación manual de habilidades técnicas y el escalado de características aseguran que los datos sean compatibles con el algoritmo k-Means, mientras que la visualización gráfica de los clusters proporciona una herramienta intuitiva para comprender los patrones subyacentes en los datos. Además, la asignación de recomendaciones de cursos específicas a cada cluster permite una toma de decisiones más informada y estratégica, lo que puede traducirse en beneficios tanto para los empleados como para la organización.

Capturas de pantalla del prototipo:

https://drive.google.com/drive/folders/12ANSiMEFK3Txx6EkYCNSh4IdF6p3q_8X?usp=drive_link