



HACKATHON

Gobierno de datos y analítica

Contenido

- Contexto General 3
 - Propósito del Reto..... 3
 - Desafío Identificado 3
 - Propuesta del Hackathon 3
- Alcance y Descripción del Reto 4
- Requisitos técnicos y funcionales..... 5
 - Requisitos Técnicos 5
 - Requisitos Funcionales 5
- Método e Instrucciones de Entrega 7
 - Método de Entrega 7
 - Instrucciones de Entrega..... 7
 - Entregables Esperados 7
- Criterios de Evaluación..... 9

Contexto General

Periferia IT Group es una compañía líder en transformación digital que ofrece soluciones tecnológicas a través de sus diversas verticales de servicio, entre las cuales destaca Staffing IT. Este servicio se especializa en proporcionar talento altamente calificado para proyectos tecnológicos, garantizando que el personal se ajuste a las necesidades técnicas y culturales de los clientes.

Propósito del Reto

El propósito de este hackathon es mejorar la eficiencia de los procesos de selección de personal en Periferia IT mediante el desarrollo de herramientas analíticas y de gobierno de datos que potencien las capacidades del equipo de reclutamiento. Actualmente, Periferia IT utiliza una plataforma de gestión de procesos de selección, donde se crean vacantes y se gestionan las postulaciones de los candidatos.

Desafío Identificado

En el contexto actual, la plataforma de gestión de procesos de selección no ofrece una funcionalidad que permita identificar a los candidatos que, a pesar de no haber sido seleccionados en un proceso específico, tienen perfiles que podrían ser ideales para otras vacantes internas de la compañía. Esto representa una pérdida de potenciales talentos y una oportunidad para aumentar la eficiencia del equipo de reclutamiento.

Propuesta del Hackathon

Desde la perspectiva de la analítica y el gobierno de datos, el reto del hackathon consiste en desarrollar una solución que permita realizar un "matching" entre los perfiles descartados recientemente y otras ofertas de trabajo disponibles en la empresa. Esta herramienta deberá proporcionar un análisis cuantitativo del grado de ajuste de un perfil descartado con una vacante abierta, basándose en la centralización y gestión adecuada de los datos de candidatos y vacantes.

La solución deberá integrar principios de gobierno de datos para asegurar la calidad y trazabilidad de la información gestionada. De esta manera, se busca no solo mejorar la eficiencia de los procesos de contratación, sino también garantizar un uso óptimo y seguro de los datos en todo el proceso.

Alcance y Descripción del Reto

Duración del Reto

Se estima que la realización del reto, dependiendo del nivel de experiencia del participante, podría tomar entre 20 a 30 horas de desarrollo.

Alcance del Reto

El reto tiene como objetivo desarrollar una solución de analítica y gobierno de datos para optimizar los procesos de selección de personal en Periferia IT. La solución debe trabajar con un conjunto de datos anonimizados que se extraerán desde la plataforma de selección de personal de la empresa. Estos datos estarán disponibles en un Blob Storage de Azure en formato Parquet y deberán ser utilizados para construir una herramienta que mejore la eficiencia del equipo de reclutamiento.

Las áreas clave del proyecto incluyen:

- **Centralización de Datos:** Integración y catalogación de los datos anonimizados en un repositorio centralizado.
- **Desarrollo de Algoritmos de Matching:** Implementación de un algoritmo de machine learning para evaluar la compatibilidad de perfiles descartados con nuevas vacantes.
- **Visualización:** Creación de dashboards interactivos que muestren el grado de afinidad de los candidatos con las vacantes disponibles.
- **Gobernanza de Datos:** Asegurar la calidad y trazabilidad de los datos utilizados.

Descripción del Reto

En el contexto actual, se identificó la necesidad de mejorar el proceso de selección mediante el aprovechamiento de los datos disponibles. El hackathon propone desarrollar una solución que:

1. **Recolecte y Centralice los Datos:** Utilizando los datos anonimizados proporcionados, centralizar la información en un sistema que permita un fácil acceso y análisis.
2. **Desarrolle un Sistema de Matching:** Implementar un sistema basado en machine learning que permita evaluar y sugerir perfiles descartados para otras vacantes disponibles en la empresa, asignando un valor de afinidad o compatibilidad a cada candidato en relación con las ofertas de empleo actuales.
3. **Visualización y Reportes:** Crear dashboards en herramientas como Power BI que permitan visualizar los resultados del matching y apoyar la toma de decisiones del equipo de reclutamiento.
4. **Gobernanza de Datos:** Incorporar medidas de gobierno de datos que aseguren la integridad y trazabilidad de la información a lo largo del proceso, cumpliendo con las normativas y mejores prácticas de gestión de datos.

El objetivo final es no solo aumentar la eficiencia del equipo de reclutamiento, sino también asegurar que los talentos potenciales no se pierdan en procesos de selección menos optimizados.

La solución propuesta deberá ser escalable, eficiente, y alineada con las necesidades estratégicas de Periferia IT.

Requisitos técnicos y funcionales

Requisitos Técnicos

1. **Plataforma de Almacenamiento Inicial**
 - **Azure Blob Storage:** Los datos anonimizados estarán disponibles en Azure Blob Storage en formato Parquet. Estos datos servirán como punto de partida para que los participantes inicien el desarrollo de la solución.
2. **Lenguajes de Programación**
 - **Python:** Recomendado para el desarrollo de scripts y algoritmos.
 - **R:** Alternativa para análisis y modelado.
 - **PySpark:** Para procesamiento distribuido si se trabaja con grandes volúmenes de datos.
3. **Frameworks y Bibliotecas**
 - **scikit-learn:** Para la implementación de algoritmos de machine learning.
 - **TensorFlow o PyTorch:** Opcionales para modelos avanzados.
 - **Pandas y NumPy:** Para manipulación de datos.
4. **Opciones de Desarrollo**
 - **Desarrollo en la Nube:** Uso de servicios en nubes como Azure, AWS, GCP. (no se incluye presupuesto para el desarrollo).
 - **Desarrollo Local:** Los participantes pueden optar por desarrollar en entornos locales configurados según sus preferencias.
5. **Herramientas de Visualización**
 - **Power BI, Looker Studio o Tableau:** Para crear dashboards interactivos que visualicen los resultados del análisis.
6. **Repositorios de Código**
 - **GitHub:** Para el almacenamiento y gestión del código, con una estructura clara y documentación adecuada.

Requisitos Funcionales

1. **Centralización de Datos**
 - **Acceso y Catalogación de Datos:** Centralizar y gestionar los datos en un sistema que permita su análisis eficiente.
2. **Desarrollo de Algoritmos de Matching**
 - **Matching de Perfiles:** Implementar un algoritmo de machine learning que evalúe la compatibilidad entre los perfiles descartados y las vacantes disponibles.
 - **Evaluación y Optimización:** Asegurar que el algoritmo sea preciso y relevante para las necesidades de selección.
3. **Visualización**

- **Dashboards Interactivos:** Crear visualizaciones que permitan al equipo de reclutamiento identificar fácilmente los candidatos más adecuados para las vacantes.
 - **Filtrado y Búsqueda:** Incluir opciones que permitan a los usuarios filtrar y buscar perfiles según diferentes criterios.
4. **Gobernanza de Datos**
- **Trazabilidad:** Asegurar se mantenga la trazabilidad de cualquier cambio realizado sobre los datos.
 - **Calidad:** Incluir escenarios de valoración de los datos que se van a usar en el proyecto.
5. **Documentación**
- **Arquitectura y Código:** Proveer documentación clara que explique la arquitectura y el código de la solución.
 - **Manual de Usuario:** Incluir un manual de uso con instrucciones y ejemplos para facilitar la adopción de la solución desarrollada.

Método e Instrucciones de Entrega

Método de Entrega

Los participantes deberán subir sus entregables a un repositorio de GitHub, el cual puede ser público o privado según la preferencia del equipo. Además, deberán enviar un correo electrónico a las direcciones henrycuasquer@periferia-it.com, jesusmorales@periferia-it.com, paulamunoz@periferia-it.com y julianmayorquin@periferia-it.com

Instrucciones de Entrega

1. **Organización del Repositorio**
 - El repositorio debe estar bien organizado, con carpetas claramente etiquetadas para código, documentación, dashboards, y el catálogo de datos.
 - Incluir un archivo README en la raíz del repositorio.
2. **Envío del Enlace**
 - Una vez finalizado el desarrollo, los equipos deben enviar el enlace del repositorio de GitHub a la dirección de correo electrónico indicada por los organizadores.
 - En el correo electrónico, se debe incluir:
 - Nombre del participante.
 - Descripción breve de la solución.
3. **Fecha Límite de Entrega**
 - Todos los entregables deben ser enviados antes del 25 de septiembre de 2024. Entregas fuera del plazo no serán consideradas.
4. **Formato de Documentación**
 - Toda la documentación técnica debe estar en formato PDF o Markdown dentro del repositorio.
5. **Revisión Previa**
 - Se recomienda que los equipos revisen cuidadosamente todo el contenido antes de la entrega para asegurar que todos los archivos estén completos y correctamente documentados.

Entregables Esperados

Los participantes del hackathon deberán entregar los siguientes elementos:

1. **Código Fuente**
 - **Repositorio en GitHub:** El código debe estar alojado en un repositorio público o privado, con una estructura clara que incluya carpetas organizadas y un archivo README con instrucciones detalladas de instalación y uso.
 - **Modularización y Buenas Prácticas:** El código debe estar modularizado, utilizando patrones de diseño y buenas prácticas de programación que aseguren su mantenibilidad y escalabilidad.
2. **Documentación Técnica**
 - **Descripción de la Arquitectura:** Un documento que explique la arquitectura de la solución, incluyendo el flujo de datos, las decisiones técnicas y las herramientas utilizadas.

- **Trazabilidad de Datos:** Descripción de cómo se gestionaron los datos, asegurando su trazabilidad.

3. Catálogo de Datos

- **Metadatos:** Información relevante sobre cada conjunto de datos, incluyendo descripciones, formatos, fechas de creación, y cualquier otra característica importante.
- **Linaje de Datos:** Un catálogo que documente todos los datos utilizados en el proyecto, detallando el origen, transformación y destino final de cada conjunto de datos, asegurando un linaje claro y completo.

4. Dashboards y Visualizaciones

- **Interactividad y Usabilidad:** Dashboards que presenten de manera clara y efectiva los resultados del análisis de matching, con opciones de filtrado y búsqueda para facilitar la toma de decisiones.
- **Visualización del Matching:** Gráficos y tablas que muestren la afinidad entre perfiles descartados y vacantes abiertas, proporcionando insights útiles para el equipo de reclutamiento.

5. Presentación Final

- **Pitch o Demostración:** Una presentación que resuma la solución desarrollada, destacando los aspectos más innovadores y efectivos. Debe incluir una demostración en vivo del funcionamiento de la solución.
- **Documentos de Soporte:** Cualquier documento adicional que respalde la presentación, como slides, notas explicativas o videos.

Criterios de Evaluación

Los proyectos presentados en el hackathon serán evaluados según los siguientes criterios:

1. Funcionalidad (20%)

- **Cumplimiento de Requisitos:** La solución debe cumplir con todos los requisitos técnicos y funcionales establecidos.
- **Matching de Perfiles:** Eficiencia y precisión del algoritmo para identificar y asignar perfiles descartados a nuevas vacantes.

2. Eficiencia (10%)

- **Rendimiento:** La solución debe ser capaz de manejar grandes volúmenes de datos sin comprometer el rendimiento.
- **Escalabilidad:** Capacidad de la solución para adaptarse a un mayor número de datos o a diferentes escenarios de uso.

3. Calidad del Código (10%)

- **Estructura y Mantenimiento:** El código debe estar bien organizado, modularizado y ser fácil de mantener.
- **Documentación:** La solución debe incluir una documentación clara y completa, explicando la arquitectura, el código y el uso de la solución.

4. Creatividad e Innovación (20%)

- **Innovaciones Adicionales:** Se valorarán las funcionalidades o mejoras adicionales implementadas que vayan más allá de los requisitos mínimos.
- **Originalidad:** La originalidad de la solución propuesta en términos de enfoque y tecnología utilizada.

5. Gobernanza de Datos (20%)

- **Trazabilidad y Gestión de Datos:** Implementación de medidas efectivas de trazabilidad y gobernanza de datos.
- **Catálogo:** Existencia de un catálogo de datos.

6. Presentación (20%)

- **Claridad y Efectividad:** La solución debe ser presentada de manera clara y comprensible, mostrando cómo resuelve el problema planteado.
- **Impacto Visual:** Calidad de los dashboards y visualizaciones, asegurando que sean intuitivos y útiles para la toma de decisiones.