**CARTA DEL PROYECTO**

**Conocimiento Del Negocio**

* **¿Quién es el cliente, en qué dominio comercial se encuentra el cliente?**

El cliente es Diagonal Colombia, es una corporación de derecho privado, sin ánimo de lucro, constituida en abril de 1950 para facilitar el aprovisionamiento de materias primas a la industria textil colombiana. En la actualidad Diagonal comercializa el 100% de la fibra de algodón consumida por la industria textil nacional e importa otros insumos como hilazas, fibras sintéticas y telas.

* **¿Qué problemas comerciales estamos tratando de abordar?**

El problema que estamos abordando es logístico ya que se quiere agrupar y clasificar el algodón según sus características basados en la producción por regiones y sitios de recibo con respecto a las solicitudes del cliente, de esta forma optimizar el almacenamiento y preparación de pedidos en la logística.

**Alcance**

* ¿Qué soluciones de ciencia de datos estamos tratando de construir?
  1. En primera instancia se pretende realizar una clusterización de las calidades del algodón basados en sus características.
  2. En segunda instancia se pretende clasificar las pacas de algodón según la calidad, basados en las solicitudes de los clientes, zonas de producción y sitios de recibo.
* ¿Qué haremos?
  1. Definición del problema.
  2. Adquisición de los datos.
  3. Origen de los datos.
  4. Entendimiento de los datos.
  5. Ambiente tecnológico.
  6. Almacenamiento de los datos.
  7. Frameworks de procesamiento de datos.
  8. Tratamiento de los datos.
  9. Ingesta de los datos.
  10. Limpieza y filtrado de los datos a través de métricas y distancias adecuadas.
  11. Identificación de outliers.
  12. Aplicación de valores y vectores propios en la identificación de proyecciones y reducción de dimensión “**PCA**”.
  13. Normas y determinantes para calcular varianzas globales.
  14. Productos internos para la identificación de asociaciones lineales entre variables.
  15. Número condición de las matrices de covarianzas.
  16. Creación de modelos
  17. Modelos de clusterización.
  18. Modelos de calificación.

1. Validación de los modelos.
2. Despliegue de modelos.

* ¿Cómo va a ser consumido por el cliente?

El cliente va a consumir los resultados de los modelos por medio de un dashboard en la plataforma de su elección.

**Personal**

* Quiénes están en este proyecto:
  + Diagonal-cluster:
    - **Lider del Proyecto**.
      * César Augusto Cardona Ortiz.
    - **PM**.
      * Carlos Enrique Salazar Escobar.
    - **Científico de datos**
      * Jesus Alberto Arcia Hernandez.
      * Mauricio Jose Jimenez Sanchez.
    - **Gerente de cuentas.**
      * Jefferson Adolfo Quiroz Fino.
  + Cliente:
    - **Administrador de datos**, personal de Diagonal Colombia.
    - **Contacto de negocio,** personal de Diagonal Colombia.

**Métrica**

* ¿Cuáles son los objetivos cualitativos?
  1. Reducir los recorridos de los montacargas en el almacenamiento y preparación de pedidos.
  2. Incrementar la capacidad de despachos diarios de vehículos.
  3. Aumentar la experiencia del cliente garantizando pedidos rápidos y oportunos.
* ¿Cuáles son las métricas cuantificables del proyecto?
  1. Reducir el 50 % los recorridos de los montacargas en el almacenamiento y preparación de pedidos.
  2. Incrementar en un 20 % la capacidad de despachos diarios de vehículos.
* Cuantifique las mejoras de los valores de las métricas que son útiles para el escenario del cliente.
  1. Disminución del tiempo de uso de los montacargas.
  2. Reducción en la preparación de los pedidos.
* ¿Cuál es el valor de referencia (actual) de la métrica?
  1. Tiempo de uso del montacarga. Aquí formula.
  2. Tiempo en la preparación de los pedidos. Aquí formula
* ¿Cómo mediremos la métrica? (por ejemplo, prueba A / B en un subconjunto especificado durante un período específico; o comparación del rendimiento después de la implementación con la línea de base)

**Plan**

* Fases (hitos), línea de tiempo, breve descripción de lo que haremos en cada fase.

Ver diagrama de Gantt



**Arquitectura**

* Datos
  + ¿Qué datos esperamos?

Esperamos archivos .csv descritos de la siguiente manera a un tiempo de 10 años:

* + - * Características de recibos de mercancías.
      * Características de despachos de mercancía.
      * Movimientos de pacas entre sitios de almacenamientos.
      * Históricos de las condiciones meteorológicas de las zonas de producción.
      * Localización histórica de las pacas en sitios de distribución.
* Movimiento de datos de local a AWS usando el componente S3 mediante boto3 y sqoop, donde vamos a trabajar sobre el siguiente esquema:
  + **Raw**, bucket de S3 donde van a estar alojados los datos crudos.
  + **Processed,** bucket de S3 donde van a estar alojados los datos procesados.
  + **Modeling,** bucket de S3 donde van a estar alojados los datos para realizar el modelado.
* **Herramientas y recursos de análisis**
  + microsoft project 2019 para el modelado del proyecto.
  + Simple Storage Service (Amazon S3) para el almacenamiento de la informacion.
  + Python para construcción de las ETL y modelado de la información.
  + Amazon Elastic MapReduce (Amazon EMR).
* **Consumo de los modelos creados en el proyecto**
  + ¿Por medio de que herramienta se va a consumir la información suministrada por los modelos?

Por medio de la fuente de datos ODBC de POWER BI utilizando AWS S3.

* + ¿Cómo utilizará el cliente los resultados del modelo para tomar decisiones?

Se montará en la página web “***https://www.diagonal.com.co/#”*** en la sección de “***HVI***” un dashboard con los filtros de agencias de recibo, agremiación y fecha. El aplicativo debe sugerir el almacenamiento de las mercancías recibidas.

* + Tubería de movimiento de datos en producción
    - Se realiza la ingesta de datos a bucket raw por medio de sqoop.
    - Luego se realizan transformaciones a los datos y esta información queda alojada en el bucket processed.
    - Se realizan transformaciones de los datos al esquema definido en producción, esta información queda alojada en el bucket modeling.
  + Diagrama de arquitectura de la solución:

**Imagen que contiene camiseta, negro

Descripción generada automáticamente**

**Comunicación**

* ¿Cómo nos mantendremos en contacto?

Se tienen proyectadas reuniones diarias para verificar el progreso de cada una de las tareas programadas.

* ¿Quiénes son las personas de contacto en ambos lados?

Por parte de Diagonal Colombia el gerente financiero Carlos Enrique Salazar Escobar que es el **PM.**