
Resumen

El problema que se desarrolla en este Trabajo de Fin de Grado es un mapeo, correspondiente al cálculo de CVA (Credit Valuation Adjustment). Esta idea es válida para ser aplicada en cualquier otro tipo de problema que requiera mapear la información para la obtención de valores relevantes, tales para llevar a cabo decisiones bancarias, con la idea de obtener un beneficio. Para este caso, el objetivo principal es soportar las pérdidas derivadas de un deterioro en la calidad crediticia de sus contrapartidas, en definitiva, conocer el riesgo de un cliente que solicita a la entidad bancaria un préstamo. A través de un análisis intensivo sobre los objetivos que se reclaman, este trabajo permitirá determinar el riesgo real en un conjunto de datos simulados.

Este problema descrito es un proceso ETL, encargado de extraer información, transformarla para cumplir objetivos y llevar a cabo la carga de los mismos. Se quiere desarrollar en un entorno empresarial, que trabaja con ingentes cantidades de datos, lo que motiva la realización de un breve estudio y análisis sobre las tecnologías Big Data, un término que se encuentra a la orden del día cuando se resuelven este tipo de problemas que pretenden llevarse a cabo en una empresa. Además, requiere conocer cómo se distribuye el trabajo y se aplican metodologías para trabajar en equipo, así como qué pasos y herramientas permiten desplegar el problema en entornos productivos.

Es por ello que gracias a mi experiencia personal trabajando en este sector, se propusieron una serie de objetivos, tales como conocer, analizar y alcanzar de forma satisfactoria una solución del problema CVA dentro del contexto de Big Data en un entorno empresarial basado en bancos, trabajando con grandes volúmenes de datos obteniendo una visión generalizada de cómo se trabajan con estos y para qué, abarcando así un segundo objetivo. Cuando se menciona a la empresa, se enlaza con un tercer objetivo que supone la idea de conocer las metodologías de trabajo existentes en proyectos de gran envergadura, permitiendo una coordinación óptima. Para ello, se dan a conocer los roles dentro de un equipo de trabajo y se da una visión generalizada de cómo será el equipo que desarrolla este problema en producción. De esta manera, cuando se resuelve el problema CVA también se abarca un cuarto objetivo que se relaciona con el nivel técnico, donde se muestran los lenguajes empleados y se describen los que han sido empleados en este trabajo. Por último, otro objetivo muy relevante consiste en obtener los conocimientos más relevantes acerca del entorno productivo anteriormente mencionado, pues se comentan los aspectos más importantes sobre los problemas reales que pueden darse en un proceso final, mostrando cómo se implanta en este ámbito y cómo de importante es el proceso de automatización para dar una continuidad a este problema, desarrollado en un entorno local que ha avanzado por las distintas fases hasta este punto.

Finalmente he implementado un script en Python para trabajar con las tablas implicadas, la unidad principal de trabajo que en Python se representa como *DataFrames* para conocer qué tipo de dato es el que almacenan y con el que se trabaja. Una vez que se llevan a cabo las operaciones que solventan el mapeo de los datos y se generan las tablas, este trabajo también utiliza tecnologías propias de Big Data como Pyspark, ofreciendo la misma solución que ofrece Python pero llevada a un entorno productivo.

Por ello para la resolución eficiente de CVA, a través de un proceso iterativo e incremental, se extrae la información, se llevan a cabo una serie de transformaciones y se finaliza con la carga de los datos. Esta solución pasa por la obtención de dos campos generados tras el mapeo sobre los que será necesaria la interpretación del cliente de la empresa para llevar a cabo la toma de decisiones más adecuada.

Contenidos

1. Bibliografia	1
-----------------	---

ndice de figuras

ndice de fragmentos de cdigo

ndice de tablas

1. Bibliografia