Trabajo Final 2020-2

Sección: Única

Logro del curso

Al finalizar el curso, el estudiante será capaz de recuperar, procesar, analizar e interpretar datos multidimensionales desde diferentes orígenes de datos y herramientas de visualización

Objetivo del Proyecto Final

Desplegar una plataforma de soporte a proyectos en ciencia de los datos

Introducción

El trabajo final es un producto software que desarrolla el ciclo de cicla de un proyecto de ciencia de datos, la elección del caso de estudio es elegido por los integrantes del grupo (máximo 2).

A partir del tema elegido en el proyecto parcial tendrá que aplicarlo en cualquier dataset publico peruano.

Especificaciones

1. **Infraestructura y Plataforma de Información (2 puntos)**

Infraestructura
Creación de 2 máquinas virtuales o contenedores en el sistema operativo de su elección.
Instalación y Configuración en Plataformas Cloud (Elegir una de ellas)
1. Microsoft Azure 2. Google Cloud 3. IBM Bluemix 4. Digital Ocean 5. AWS 6. Otros

Configuración de Plataforma en VPS
Web, Base de datos, Procesamiento, Visualización de datos
• Instalación y configuración de Apache2, ssh, Mysql, Postgresql, SQLite (opcional) • Instalación y configuración de Lenguajes de Programación R (VM_Procesamiento) • Configuración de túnel ssh

2. **Ciclo de vida de ciencia de los datos**

Fase	Detalle
Recolección de los Datos **(1 punto)**	Selección de 2 Fuentes de datos como: • CSV, TSV, EXCEL

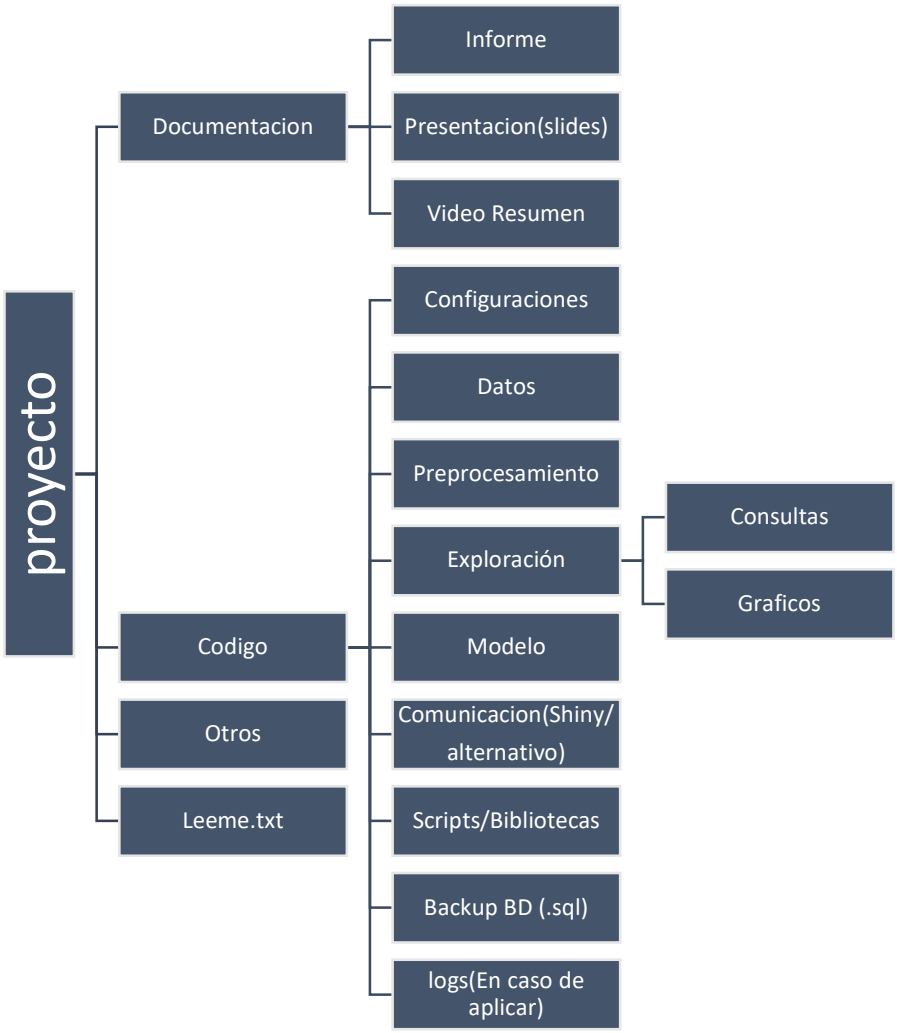
	• JSON, XML • BD: Mysql, SQLite, Postgres, Neo4J • Recolección de datos por web scraping (si se requiere) • Especificación del Conjunto de datos
Modelado de datos estructurados **(2 punto)**	• Modelado • Implementación • Poblado • Serialización
Transformación y consultas exploratorias **(3 punto)**	Transformación (10 consultas por integrante) • Selección • Unión • División • Filtrado Estadística descriptiva ○ Media ○ Máximo ○ Mínimo ○ Cuartiles ○ Percentiles
Preparación de los datos **(2 puntos)**	• Muestreo • Normalización • Imputación • Eliminación de valores anómalos • Eliminación de outliers
Exploración Visual de datos **(3 puntos)**	Dependiendo del caso de estudio aplicar los diagramas de • Diagramas de dispersión • Diagramas de barras • Diagramas de cajas • Diagramas de serie de tiempo (10 diagramas por integrantes de complejida incremental)
Modelo **(4 puntos)**	• Regresión lineal • Regresión no lineal(polinomial) • SVM • PCA • KNN • Agrupamiento (K-means, K-Median) • Reglas de Asociación • Series de Tiempo (seleccionar 2 modelos por integrante)
Exportación y Comunicación **(1 punto)**	Exportación de datos • CSV, TSV, EXCEL • JSON, XML • BD: Mysql, SQLite, etc

3. Funcionalidades adicionales **(2 puntos, seleccionar 1 por integrante)**
    - Envío de Correo
    - Generación de Informe automático
    - Creación de un servicio API Rest
    - Creación de una Biblioteca Propia
4. Informe **(1 punto)**
    - Descripción del caso de estudio
    - Procedimiento
    - Conclusiones
5. Presentación **(1 punto)**
    - Presentación

Consideraciones en el proyecto

- El proyecto debe estar alojado en un repositorio asignado en github, conteniendo la estructura de directorios indicada.
- En la fase de recolección debe contener al menos 2 fuentes de datos distintos.
- En la fase de transformación, exploración y visualización debe contemplar al menos 10 por integrante consultas de complejidad incremental y variada por cada integrante de grupo.
- En la fase de modelo distinto aplicar dos modelos por cada integrante de grupo, puede utilizar un modelo no necesariamente de los que se encuentran listado si es fundamentado.
- De las funcionalidades adicionales listadas se elige uno por cada integrante de grupo.
- El informe se realizará en formato Markdown o Latex.
- **Es obligatorio que la presentación del proyecto sea en aplicación web en su propia plataforma, siguiendo el ciclo de vida planteado de manera integrada,** la no implementación disminuirá en 4 puntos a su calificación final obtenida.
- Un integrante del grupo deberá cargar el entregable solicitado en el github respectivo, no se acepta los envíos por correo electrónico fuera de la fecha límite. La calificación se otorga solamente a los integrantes que **realiza dicho envío y expone**.
- Cada integrante del grupo debe conocer la **funcionalidad por completo**, no se considerará los argumentos similares a esto: "…en esta parte me ayudaron", "….mi compañero lo hizo", **la calificación de la exposición es individual**.
- Si se detecta que se presenta un proyecto anterior se calificará con cero(0).
- Cualquier duda respecto al trabajo o al alcance se formaliza por correo con mínimo de un día de antelación o por el delegado.

Estructura de Directorios de Proyecto



Rubrica de Calificación de Trabajo Final

- Se evaluará según puntaje citado en cada especificación y se asignará el **puntaje total** si cumple todo lo especificado en cada punto.

Rubrica de Calificación de Exposición

- El trabajo Final tiene 2 Rúbricas: Competencia General (Técnica) y Comunicación Oral
- Como la competencia general del curso es Comunicación Oral, para la exposición es **obligatorio** llevar consigo una Infografía, PPT, Video Presentación Resumen del trabajo (máx 3 minutos).
- La calificación es por integrante, se adjunta la rúbrica de calificación de Comunicación Oral.

Rúbrica Actividades Blanda-Comunicación Oral – Rúbrica de Exposición			
Criterios	**Nivel Deficiente**	**Nivel Satisfactorio**	**Nivel Optimo**
Capacidad de explicación	Evalúa un problema planteando una explicación.	Evalúa críticamente un problema, lo plantea, describe y aclara.	Evalúa críticamente un problema, lo plantea con claridad y lo explica exhaustivamente, proporcionando información suficiente.
	(0-1 puntos)	**(2-3 puntos)**	**(4 puntos)**
Evaluación y cuestionamiento de la información	Desarrolla análisis y síntesis organizados, presentando la información sin evaluarla. Asume los puntos de vista de los expertos sin cuestionarlos.	Desarrolla análisis y síntesis organizados, presentando la información evaluada parcialmente. Cuestiona de modo parcial y general los puntos de vista de los expertos.	Evalúa la información para desarrollar análisis y síntesis coherentes. Cuestiona de modo general los puntos de vista de los expertos.
	(0-1 puntos)	**(2-3 puntos)**	**(4 puntos)**
Análisis del contexto y los supuestos	Identifica los supuestos ajenos, pero todavía le cuesta identificar los propios. Identifica algunos contextos relevantes cuando plantea un punto de vista.	Identifica los supuestos propios y ajenos toma en cuenta algunos contextos relevantes cuando plantea un punto de vista.	Analiza de manera general los supuestos propios y ajenos, y toma en cuenta la relevancia del contexto cuando plantea un punto de vista.
	(0-1 puntos)	**(2-3 puntos)**	**(4 puntos)**
Planteamiento y sustento de una postura	Plantea una postura presentando un sustento de forma obvia y simplificada.	Plantea una postura presentando un sustento reconociendo e incluyendo diferentes aspectos del asunto tratado.	Plantea una posición entrando en detalles que complejizan y enriquecen el asunto tratado, y se sustenta de modo consistente.
	(0-1 puntos)	**(2-3 puntos)**	**(4 puntos)**
Formulación de conclusiones	Las conclusiones están ligadas de manera poco consistente a la información discutida (poco pertinentes).	Las conclusiones están lógicamente ligadas a la información seleccionada, pero contienen sesgos.	Las conclusiones están lógicamente ligadas a información pertinente e incluyen puntos de vista distintos.
	(0-1 puntos)	**(2-3 puntos)**	**(4 puntos)**