**Plan de Trabajo para el Grupo de Estudio**

**1. Unir las bases de datos de secretaría de salud.**

**2. Listar El número variables objetivo**

**a. Número de casos**

**b. Incidencia de casos.**

**c. Las variables entomológicas:**

**d. Otra variable objetivo es el tipo de dengue, con signos de alarma 220 o sin signos de alarma 210.**

**3. Organizar datos de la secretaría de salud.**

**a. Eliminar los datos de datos que no sean de Caucasia. Resaltar para Johana los casos que no son de Caucasia.**

**b. Arreglar lo de barrio y dirección.**

**c. Organizar cronológicamente (semana epidemiológica y el año) la base de datos de acuerdo a la semana epidemiológica por año.**

**d. listar las veredas que no pertenecen a Caucasia para eliminar estas observaciones.**

**e. Preguntar criterio de inclusión para Guarumo, Campoalegre, La Apartada, Jardín. Determinemos el número de casos en esta condición.**

**f. Organizar la columna w, en cuanto a los nombres de los barrios para colaborar a Johana. Con base en la columna X alimentar la W.**

**4. Consultar si en cuanto a las variables meteorológicas, consideraremos solo a Caucasia. Y las veredas las determinaremos de acuerdo a la cercanía a Caucasia.**

**5. Hay en la base de datos casos de Cereté, Yarumal, Medellín, San Pues, Apartadó, BuenaVista, Carolina, Concordia, Itaguí, Necoclí, Remedios, TierraAlta. Estos los resaltaremos con un color para cada uno. El atributo fue municipio de residencia.**

**6. Tomemos de la base de datos de la Secretaría de salud**

**a. los números de casos.**

**b. Semana epidemiológica asociada al año coorrespondiente.**

**c. Edad**

**d. Género.**

**e. Localidad.**

**7. Hay tres atributos de de fecha, requerimos saber qué indica cada fecha y cuál vamos**

**a usar.**

**8. Queremos saber si uni\_med es el atributo o variable que describe el número de casos.**

**9. Obtener los datos meteorológicos desde 2010. para juntarlos a la base de tos de la**

**secretaría de salud y luego hacer el análisis de correlaciones.**

**Primero unir las bases de datos de Secretaría de Salud.**

**Chequear que no se vallan a duplicar los datos.**

**Luego organizar la columna w.**

**Graficar casos vs semana epidemiológica por año..**

**Los datos los trabajaremos con fecha de frecuencia semanal.**

**Qué es semana epidemiológica**

**Debemos averiguar la ubicación de cada estación.**

**Buscamos una asociación espacial más precisa.**

**Objetivo General:**

**Implementar y comparar modelos ARIMA, SVM y Random Forest en la clasificación de casos de infección por una enfermedad transmitida por vectores, utilizando series de tiempo de datos meteorológicos recolectados desde 2007 hasta 2024.**

**Duración del Proyecto:\*\***

**16 semanas (4 meses)**

**#### \*\*Integrantes:\*\***

**Tres o cuatro investigadores**

**---**

**\*\*Semana 1: Introducción y Planificación\*\***

**- \*\*Objetivo:\*\* Establecer el marco de trabajo, asignar roles y responsabilidades, y definir los objetivos específicos.**

**- \*\*Actividades:\*\***

**- Reunión inicial para discutir el objetivo del proyecto.**

**- Definición de roles: un responsable de la recolección de datos, uno de preprocesamiento de datos, uno de modelado y análisis, y uno de documentación (si hay cuatro miembros).**

**- Establecer el cronograma detallado.**

**- Identificar y obtener acceso a las fuentes de datos (IDEAM, Antioquia Mira su Cielo, Fedearroz).**

**---**

**\*\*Semana 2-4: Recolección de Datos\*\***

**- \*\*Objetivo:\*\* Obtener y organizar los datos meteorológicos y de contagios de 2007 a 2024.**

**- \*\*Actividades:\*\***

**- Contactar a las instituciones para la obtención de los datos.**

**- Descargar y organizar los datos en una estructura adecuada.**

**- Almacenar los datos en un repositorio compartido (por ejemplo, Google Drive, Dropbox, GitHub).**

**---**

**### \*\*Semana 5-6: Preprocesamiento de Datos\*\***

**- \*\*Objetivo:\*\* Limpiar y preparar los datos para el análisis.**

**- \*\*Actividades:\*\***

**- Limpieza de datos: manejar valores faltantes, corrección de errores, unificación de formatos.**

**- Empalmar los datos meteorológicos con los datos de contagios.**

**- Generar nuevas variables si es necesario (por ejemplo, medias móviles, desviaciones estándar).**

**- Visualización preliminar de los datos para entender patrones y tendencias.**

**---**

**### \*\*Semana 7-8: Exploración y Análisis de Datos\*\***

**- \*\*Objetivo:\*\* Realizar un análisis exploratorio de los datos para identificar patrones y correlaciones.**

**- \*\*Actividades:\*\***

**- Análisis descriptivo de las series de tiempo.**

**- Identificación de correlaciones entre variables meteorológicas y casos de infección.**

**- Uso de gráficos y visualizaciones para representar los datos.**

**---**

**### \*\*Semana 9-10: Implementación de Modelos ARIMA\*\***

**- \*\*Objetivo:\*\* Desarrollar y evaluar el modelo ARIMA para la predicción de casos de infección.**

**- \*\*Actividades:\*\***

**- Estudio y revisión de la teoría del modelo ARIMA.**

**- Implementación del modelo ARIMA utilizando un entorno de programación (Python, R).**

**- Evaluación del modelo mediante técnicas de validación cruzada.**

**- Ajuste y optimización del modelo.**

**---**

**### \*\*Semana 11-12: Implementación de Modelos SVM\*\***

**- \*\*Objetivo:\*\* Desarrollar y evaluar el modelo SVM para la predicción de casos de infección.**

**- \*\*Actividades:\*\***

**- Estudio y revisión de la teoría del modelo SVM.**

**- Implementación del modelo SVM utilizando un entorno de programación (Python, R).**

**- Evaluación del modelo mediante técnicas de validación cruzada.**

**- Ajuste y optimización del modelo.**

**---**

**### \*\*Semana 13-14: Implementación de Modelos Random Forest\*\***

**- \*\*Objetivo:\*\* Desarrollar y evaluar el modelo Random Forest para la predicción de casos de infección.**

**- \*\*Actividades:\*\***

**- Estudio y revisión de la teoría del modelo Random Forest.**

**- Implementación del modelo Random Forest utilizando un entorno de programación (Python, R).**

**- Evaluación del modelo mediante técnicas de validación cruzada.**

**- Ajuste y optimización del modelo.**

**---**

**### \*\*Semana 15: Comparación y Análisis de Resultados\*\***

**- \*\*Objetivo:\*\* Comparar los tres modelos en términos de precisión, recall, F1-score y otras métricas relevantes.**

**- \*\*Actividades:\*\***

**- Recopilación de resultados de los tres modelos.**

**- Comparación de las métricas de desempeño.**

**- Análisis de ventajas y desventajas de cada modelo.**

**---**

**Semana 16: Documentación y Presentación de Resultados**

**- \*\*Objetivo:\*\* Documentar todo el proceso y presentar los resultados finales.**

**- \*\*Actividades:\*\***

**- Redacción de un informe detallado con metodología, resultados y conclusiones.**

**- Preparación de una presentación para compartir los hallazgos con otros colegas o en un foro académico.**

**- Discusión sobre posibles trabajos futuros y mejoras al proyecto.**

**---**

**Herramientas y Tecnologías**

**- Programación: Python (bibliotecas: pandas, scikit-learn, statsmodels), R**

**- Entornos de trabajo: Jupyter Notebook.**

**- Visualización: Matplotlib, Seaborn.**

**- \*\*Almacenamiento de datos:\*\* Google Drive, Dropbox, GitHub**

**Consideraciones Finales**

**Es crucial mantener una comunicación constante entre los miembros del equipo, realizar reuniones semanales para revisar el progreso y ajustar el plan según sea necesario. Además, documentar cada etapa del proceso ayudará a asegurar la reproducibilidad y claridad del proyecto.**