

## Proyecto en Grupo – NLP - Clasificación Automática de Tickets

**Materia:** Machine Learning

**Profesor:** Jheser Guzman Ph.D.

**Fecha de Entrega:** Sabado Feb 25, 2023 hasta las 23:59

### Formato Entrega:

- Enviar email al profesor indicando el URL del repositorio de GitHub.
- Repositorio debe incluir: Dataset, Readme.md y Jupiter Notebook.

### Rúbrica de Evaluación

Críterio	(%)	Expectativas Esperadas
<b>Data Reading / Data Understanding</b>	5	El alumno ha importado los datos .json y ha creado el dataframe a partir de ellos. El alumno también tiene que leer los datos y hacer el análisis inicial a partir de ellos.
<b>Data Cleaning</b>	10	El alumno tiene que realizar las operaciones de limpieza de datos, como filtrar texto, eliminar valores faltantes y cambiar el nombre de los encabezados de las columnas.
<b>Data Preprocessing</b>	10	Los pasos de preprocesamiento de datos se realizan mediante operaciones estadísticas como Lematization y POS tagging.
<b>Data Visualización</b>	10	El alumno ha realizado los pasos de visualización para encontrar la distribución de conteo de palabras y la distribución de n-gramas.
<b>Feature Extraction</b>	5	Las características se extraen de los datos utilizando Tf-Idf.
<b>Topic Modelling</b>	25	El alumno ha utilizado Tópico Modeling para hacer lo siguiente: <ul style="list-style-type: none"> <li>• Establezca el mejor número de tópicos en 5.</li> <li>• Aplicar el mejor número para crear grupos de palabras (tópicos).</li> <li>• Inspeccionar y validar la corrección de cada grupo para cada queja.</li> <li>• Asignar los grupos a temas/nombres de grupos de tópicos.</li> </ul>
<b>Model Building</b>	20	Preparé los datos de Train y Test utilizando las etiquetas recibidas de NMF o LDA. El alumno debe aplicar al menos 2 de estos modelos en el conjunto de datos creado: <ul style="list-style-type: none"> <li>• Regresión logística</li> <li>• Árbol de decisión</li> <li>• Bosque aleatorio</li> <li>• Naïve Bayes (opcional)</li> </ul> Una vez que se construyen los modelos, el alumno debe evaluar su desempeño utilizando métricas de evaluación relevantes para seleccionar el mejor modelo.
<b>Model Inference</b>	5	El alumno ha aplicado el mejor modelo para predecir un texto personalizado y ver su rendimiento.
<b>Code readability and conciseness</b>	10	El alumno ha hecho lo siguiente: <ul style="list-style-type: none"> <li>• Los comentarios apropiados se escriben donde corresponde.</li> <li>• En general, el código es legible y tiene sangrías apropiadas.</li> </ul>