# Room Occupancy Stimation

Daniel Alejandro Yepes Mesa, --------, ----------

*Dpto. of Systems Engineering and Computer Science Universidad de Antioquia*

**Resumen**

# ´Indice

1. [Introducción](#_bookmark0)
2. Descripción
3. [Estado del arte](#_bookmark2)
4. Entrenamiento y evaluación de modelos
   1. configuración experimental
   2. Resultado de el entrenamiento de modelos
5. [reducción](#_bookmark6) de dimensión
   1. Selección de características
   2. [extracción](#_bookmark8) de características
6. [Evaluación](#_bookmark9) 4

# introducción

Cada equipo de trabajo debe seleccionar un problema que pueda ser abordado a partir de

las t´ecnicas y aproximaciones vistas en el cur- so de Modelos y Simulacio´n II. Los problemas abordados pueden ser propuestos por el gru- po de estudiantes, o seleccionado a partir del listado que ser´a compartido por el profesor del curso. **S´olo se permite que un pro- yecto/dataset sea desarrollado por un equipo de trabajo**. El listado de proyectos sugeridos puede ser consultado en el siguiente [enlace](https://docs.google.com/spreadsheets/d/1lm_159ZMxZ5CX8ueLQRIS9_IQNcUFvNHoZhM5vKvNx0/edit?usp=sharing).

Cada equipo de trabajo debe crear un repo- sitorio GitHub en el cual se debe alojar un reporte del proyecto (en formato pdf) y, al menos, un notebook en el que se puedan re- producir los resultados incluidos en el reporte. El readme del repositorio debe dar indicacio- nes claras de co´mo interactuar correctamente con el repositorio. Para el reporte se debe usar la plantilla [IEEE](https://journals.ieeeauthorcenter.ieee.org/create-your-ieee-journal-article/authoring-tools-and-templates/tools-for-ieee-authors/ieee-article-templates/) para los art´ıculos publicados en Transactions. Se recomienda usar un editor de texto LATEX, en particular [Overleaf](https://www.overleaf.com/). Tanto el reporte como el/los notebooks deben con- tener las secciones descritas a continuacio´n.

# Descripci´on del problema (20 %)

La primera parte del informe esta´ destinada a realizar una descripcio´n del problema, hacer

un an´alisis exploratorio de los datos que permita comprender co´mo est´a compuesta la base de

datos y a definir la aproximacio´n que desde el punto de vista de *Machine Learning - ML* sera´ seguida por parte del equipo para la solucio´n del problema. Todos los informes deben contener:

1. Una descripci´on clara del contexto del problema, donde se refleje la utilidad de desarrollar una solucio´n basada en ML para dicho problema.
2. Una descripcio´n de la composicio´n de la base de datos incluyendo (pero no li- mita´ndose) el nu´mero de muestras, el

caso, la estrategia de imputacio´n de da- tos seguida. Adema´s, se debe describir claramente el tipo de codificaci´on usado para las diferentes variables del proble- ma.

1. El tipo de configuracio´n o paradigma de aprendizaje que el equipo de trabajo de-

nu´mero de variables, su significado, la

cidio´

como apropiado para el abordaje

existencia de datos faltantes y si es del del problema y su justificaci´on.

# Estado del arte (10 %)

El prop´osito fundamental de esta fase del proyecto es que los grupos de trabajo revisen literatura especializada del campo de ML en la que se presenten soluciones a problemas similares al descrito en la seccio´n [2](#_bookmark1).

**Artículo 1: *Room Occupancy Prediction: Exploring the Power of Machine Learning and Temporal Insights* [1]**

* **Paradigma de aprendizaje:** Supervisado, clasificación multiclase.
* **Técnicas usadas:** Regresión logística, LDA, SVM, Random Forest, XGBoost, LightGBM, MLP. Validación: Validación cruzada de 5 pliegues, ajuste de hiperparámetros.
* **Métricas:** Precisión, AUC ponderado, análisis SHAP.
* **Resultados:** Random Forest alcanzó 98.7% de precisión. Las variables de luz fueron las más importantes según SHAP.

**Artículo 2: *Ubiquitous Multi-Occupant Detection in Smart Environments* [2]**

* **Paradigma de aprendizaje:** Supervisado, secuencial.
* **Técnicas usadas:** BiGRU, LSTM, MLP, SVM, QDA, k-NN.
* **Validación:** Separación entrenamiento/prueba, con ventanas deslizantes.
* **Métricas:** Precisión, F1-score, precisión, recall.
* **Resultados:** BiGRU logró F1-score superior al 90% en la detección de múltiples ocupantes.

# Entrenamiento y Evaluacio´n de los Modelos (30 %)

En esta seccio´n se debe describir el disen˜o experimental y los resultados obtenidos durante las simulaciones.

## Configuracio´n experimen- tal

1. Describa la metodolog´ıa de validacio´n escogida para entrenar y evaluar los mo- delos de ML. En aso de requerir etapas adicionales, como la utilizacio´n de t´ecni- cas de sub o sobre muestreo, la forma en la fueron usadas debe estar descrita en este apartado.
2. Incluya una tabla en la que especifique el conjunto de hiperpara´metros que fueron analizados durante el proceso de evalua- cio´n de cada modelo, junto con la des- cripcio´n de la malla de valores usada pa- ra cada uno. Cada proyecto debe compa- rar al menos 5 modelos de aprendizaje, entre los que se deben incluir:

 Un modelo para´metrico basado en funciones discriminantes gausianas o en regresio´n (lineal o log´ıstica segu´n se requiera).

 Un modelo no para´metrico

 Un modelo basado en el ensamble de ´arboles de decisi´on

 Una red neuronal artificial

 Una ma´quina de vectores de sopor- te

1. Describa cu´ales sera´n las m´etricas de desempen˜o que se usara`n para evaluar el sistema y justifique su utilizacio´n.

## 4.2. Resultados del entrena- miento de Modelos

1. Presente los resultados de la experimen- tacio´n para cada uno de los modelos eva- luados, en los que se evidencie el efecto de los hiperpar´ametros en el desempen˜o del modelo. Haga uso de tablas y gr´afi- cas que ayuden al lector a comprender de manera clara los resultados. **TODAS** las figuras y tablas deben ser introduci- das en el texto del informe y contener un *caption* que la describa. Todas las figuras deben incluir el nombre de los ejes para que el lector pueda comprender la informaci´on que se le presenta. Las medidas de desempen˜o deben incluir los correspondientes intervalos de confianza estimados durante la fase de validaci´on. **Recuerde incluir resultados de en- trenamiento, validaci´on y test.**

# Reducci´on de dimensio´n (20 %)

En esta fase del proyecto se deben evaluar diferentes t´ecnicas de reduccio´n de dimensio´n y establecer si es posible reducir la complejidad del modelo final.

## Selecci´on de caracter´ısticas

1. Realice un ana´lisis individual de cada una de las caracter´ısticas, a partir de medidas de correlacio´n y o ´ındice que refleje la capacidad discriminativa de las variables (segu´n sea el caso). Identifique de acuerdo con este an´alisis, ¿si existen y cu´ales son las caracter´ısticas candida- tas a ser eliminadas?
2. Realice seleccio´n de caracter´ısticas por el m´etodo de bu´squeda secuencial ascen-

dente o descendente y evalu´e el subcon- junto de caracter´ısticas encontrado, en los 2 mejores modelos predictivos encon- trados en la secci´on [4](#_bookmark3). Para realizar este punto cada grupo debe decidir la fun- cio´n criterio a usar en el algoritmo de se- leccio´n, justificando su decisio´n. Incluya una tabla con los resultados, indi-cando el porcentaje de reduccio´n alcanzado y las dema´s medidas de desempen˜o. Re- cuerde incluir en el informe cua´l fue el criterio de seleccio´n usado y porqu´e.

## Extracci´on de caracter´ısti- cas

1. Use el m´etodo PCA para encontrar un conjunto de componentes principales in- ferior al nu´mero de variables original y evalu´e nuevamente en los 2 mejores mo- delos de prediccio´n encontrados en la seccio´n [4](#_bookmark3). Cada grupo debe decidir el criterio para seleccionar el nu´mero de

componentes principales justificando su decisio´n. Incluya una tabla con los resul- tados, indi- cando el porcentaje de re- duccio´n alcanzado y las dema´s medidas de desempen˜o.

1. Incluya una seccio´n de discusi´on y con- clusiones sobre la evaluaci´on completa de la soluci´on desarrollada, e incluya una comparacio´n de sus resultados con los descritos en la secci´on [3](#_bookmark2).

# Bibliografía

# [1] S. Mao, Y. Yuan, Y. Li, Z. Wang, Y. Yao y Y. Kang, "Room Occupancy Prediction: Exploring the Power of Machine Learning and Temporal Insights," *American Journal of Applied Mathematics and Statistics*, vol. 12, no. 1, pp. 1–9, 2024. [En línea]. Disponible en: <https://pubs.sciepub.com/ajams/12/1/1/index.html>

# [2] D. Fährmann, F. Boutros, P. Kubon, F. Kirchbuchner, A. Kuijper y N. Damer, "Ubiquitous Multi-Occupant Detection in Smart Environments," *Neural Computing and Applications*, vol. 36, no. 6, pp. 2941–2960, 2024. [En línea]. Disponible en: <https://link.springer.com/article/10.1007/s00521-023-09162-z>

# Evaluaci´on (20 %)

Para la evaluacio´n final del proyecto se tendr´an en cuenta los siguientes criterios:

Alcance de los objetivos de cada una de las secciones anteriores.

No superar las 10 pa´ginas en el reporte.

Orden y claridad del repositorio GitHub para la reproduccio´n de los resultados.

Sustentacio´n: cada grupo debe preparar un video de 10 minutos en el que se pre- senten un resumen del proyecto. El vi-

deo sera´ proyectado en una sesio´n de cla- se y los estudiantes deber´an responder a las preguntas del publico. La capacidad de cada integrante del grupo de respon- der a las preguntas sera´ evaluada de ma- nera individual.

**Nota**: La seccio´n [2](#_bookmark1) y la mitad de la seccio´n [3](#_bookmark2) sera´n evaluadas en la primera entrega de avance del proyecto.