COLEGIO DE CIENCIAS E INGENIERÍAS

INGENIERÍA EN CIENCIAS DE LA COMPUTACIÓN

Planificación para el Desarrollo del Proyecto Integrador

Tutor: Ricardo Flores Moyano

Autor: John Ochoa Abad

Quito – Ecuador

2026

1. Título del Proyecto:

Arquitectura Mixture of Experts para Clasificación de Ataques en Datasets con Desbalance Extremo de Clases

2. Relevancia y Justificación:

Dentro del área de la Inteligencia Artificial, la base del desarrollo de los modelos modernos es el disponer de los datos necesarios sobre los cuales desarrollar modelos que cumplan con un propósito deseado. Por ende, la base de cualquier modelo parte del manejo, la calidad y la correcta obtención de los datos disponibles bajo un escenario determinado.

En este contexto, se han desarrollado técnicas para el manejo de datos que parten de la recolección mediante métodos como el uso de logs, monitoreos en tiempo real, sensores, web scraping, entre otros. Posterior a la obtención de la mayor cantidad de datos de buena calidad posible, es necesario proceder con la limpieza de estos y con su respectivo preprocesamiento a través de técnicas como el escalamiento, la normalización y/o el encoding. Ulteriormente, se procede a realizar feature engineering mediante métodos matemáticos de selección de características, tales como los métodos filter, wrapper o embedded, o mediante el uso de técnicas como PCA.

Bajo dicho contexto, es importante que los datos no sufran de problemas como desbalanceo, sesgo, data leakage o irrelevancia de características. De todos estos problemas, uno de vital importancia es el desbalanceo de clases, considerando que, a pesar de todos los esfuerzos por obtener datos de la mayor calidad posible, en muchas circunstancias no es factible contar con suficientes muestras para todas las categorías en problemas de clasificación.

Ante este escenario, es necesario emplear técnicas como undersampling y oversampling. Por un lado, el undersampling incurre en la pérdida de información como parte del trade-off al reducir muestras de la clase mayoritaria, mientras que el oversampling introduce técnicas más avanzadas como SMOTE, que se basa en la interpolación de datos de clases minoritarias para generar nuevos casos sintéticos, o ADASYN, que genera muestras sintéticas considerando la densidad local de los vecinos de la clase minoritaria.

No obstante, en muchos casos estas estrategias a nivel de datos no resultan suficientes. En tales situaciones, se hace necesario abordar el problema desde la ingeniería del modelo. Para ello, se pueden emplear modelos que manejan de mejor manera los datos desbalanceados, como los modelos basados en árboles de decisión, por ejemplo, Random Forest o XGBoost, los cuales tienden a ser más robustos frente al desbalance. Sin embargo, ha surgido en meses recientes el estudio de una potencial alternativa aún más potente, que es el uso de arquitecturas basadas en Mixture of Experts, las cuales se fundamentan en la idea de utilizar el mejor modelo o experto para predecir cada entrada específica, mediante el uso de una red de control (gate). De este modo, no todos los modelos contribuyen de igual manera, sino que el gate decide dinámicamente qué experto posee mayor conocimiento para ese caso particular independientemente de que la clase asociada sea mayoritaria o no.

Adicionalmente, es necesario considerar en problemas de clasificación con datos desbalanceados de igual manera han surgido estudios en torno a los beneficios del uso de la Focal Loss, una función de pérdida la cual introduce un factor de enfoque que reduce la contribución de los ejemplos bien clasificados y enfatiza aquellos ejemplos difíciles, que suelen corresponder a las clases minoritarias. De esta manera, la Focal Loss permite que el modelo concentre su aprendizaje en aquellos casos, mejorando métricas como el recall y el F1-score en las clases minoritarias. La combinación de arquitecturas Mixture of Experts con funciones de pérdida como Focal Loss resulta particularmente atractiva, ya que permite la especialización de los modelos con un entrenamiento más efectivo en presencia de desbalance severo por ende es importante seguir investigando sobre esta línea para aportar valor al potencial uso imperativo de está técnica para el manejo de datos desbalanceados.

Finalmente, la relevancia se ve aún más reforzada al considerar el dominio de aplicación: la clasificación de ataques cibernéticos. Una clasificación precisa y robusta de este tipo de ataques es sustancial, ya que permite implementar mecanismos de defensa como firewalls, sistemas de detección y prevención de intrusiones (IDS/IPS) y filtros inteligentes de tráfico que permitan bloquear ataques de forma oportuna. En un mundo cada vez más interconectado y dependiente de sistemas digitales, el desarrollo de modelos capaces de manejar datos desbalanceados de manera efectiva para la detección de ataques cibernéticos constituye una contribución relevante y necesaria ya que, en definitiva, si se puede predecir se puede evitar y por eso es tan importante el trabajar sobre el presente tema.

3. Objetivos

a. Generales

Diseñar, implementar y evaluar una arquitectura Mixture of Experts (MoE) para la clasificación multiclase de ataques cibernéticos en un dataset de detección de intrusiones altamente desbalanceado (CICIDS2017), con el objetivo de mejorar el desempeño en clases minoritarias, particularmente en términos de recall y F1-score por clase, en comparación con modelos baseline configurados mediante un proceso controlado de selección de hiperparámetros.

b. Específicos

- Analizar y preparar para entrenamiento de un modelo IA de clasificación un dataset de detección de intrusiones multiclase con desbalance severo (como CICIDS2017), realizando procesos de limpieza, normalización, codificación de variables, partición train/validation/test sin data leakage y análisis estadístico del desbalance de clases.
- Implementar y evaluar modelos base no basados en Mixture of Experts, incluyendo al menos dos clasificadores estándar (red neuronal MLP, XGBoost u otros) con el objetivo de establecer un punto de comparación inicial en términos de métricas globales y por clase.
- Aplicar y comparar estrategias clásicas para el manejo del desbalance de clases, tales como: uso de funciones de pérdida como Focal Loss, Feature Engineering y una técnica de oversampling como SMOTE analizando sus efectos sobre el desempeño del modelo y los posibles riesgos asociados.
- Diseñar e implementar una arquitectura Mixture of Experts, compuesta por: un mecanismo entrenado de gating para manejo correcto del enrutamiento, un conjunto de expertos especializados y la integración de una función de pérdida focalizada en mejorar la atención sobre las clases minoritarias para clasificar correctamente las intrusiones del dataset a través de un modelo de clasificación multiclase con foco en clases minoritarias.
- Evaluar el desempeño de la arquitectura MoE propuesta, utilizando métricas apropiadas para escenarios desbalanceados como recall por clase y F1-score macro a manera de efectuar comparaciones directas de mejoría respecto a los modelos baseline.
- Garantizar la reproducibilidad experimental del proyecto, mediante la implementación de scripts, control de semillas aleatorias, registro de configuraciones, logs de entrenamiento y documentación clara de la arquitectura completa.

4. Estado del Arte

Dentro del área de las Ciencias de la Computación se ha realizado una investigación referente a determinar qué proyectos de tinte similar se han llevado a cabo recientemente respecto al tema: Arquitectura Mixture of Experts para Clasificación de Ataques en Datasets con Desbalance Extremo de Clases. Para esto, a continuación, se resumen las investigaciones encontradas y se comenta como el presente proyecto aporta respecto a los avances potenciales en el área:

- **Modelos tradicionales de Machine Learning para detección de intrusiones:** Investigaciones iniciales emplearon clasificadores como SVM, K-NN, Random Forest, XGBoost y MLP sobre datasets clásicos como UNSW-NB15, NSL-KDD y CICIDS2017. Si bien estos enfoques lograron resultados aceptables, presentan limitaciones significativas de generalización y escalabilidad, especialmente en escenarios con desbalance extremo [1]. **Aporte del proyecto:** Se supera este enfoque al emplear arquitectura Mixture of Experts condicionada a los inputs, capaces de adaptarse dinámicamente al tipo de tráfico y buscando validar el aporte al manejo de clases minoritarias.
- **Uso de CNNs mediante transformación de tráfico de red a representaciones tipo imagen:** Estudios recientes convierten flujos de red en matrices 2D para entrenar CNNs o aprovechar modelos preentrenados como VGG, ResNet o EfficientNet, obteniendo mejoras notables en precisión. Sin embargo, estos modelos utilizan capas densas estáticas, activando todos los parámetros en cada inferencia, lo cual incrementa el costo computacional [2]. **Aporte del proyecto:** Se busca mediante el presente proyecto demostrar que el delegar la decisión final a un Mixture of Experts, reduciendo inferencia innecesaria y mejorando eficiencia en el costo computacional del algoritmo de clasificación de intrusiones
- **Modelos secuenciales basados en RNN, LSTM y GRU con mecanismos de atención:** Diversos trabajos explotan la naturaleza temporal del tráfico mediante LSTM/GRU y mecanismos de atención basándose en las investigaciones recientes de Google logrando buenos resultados en clasificación de ataques con secuencia temporal [3]. No obstante, estos modelos tienden a ser costosos en entrenamiento y sensibles a datasets pequeños o altamente desbalanceados.

  Aporte del proyecto: El Mixture of Experts que se planea desarrollar busca demostrar como el especializar expertos en distintos patrones de tráfico sin depender exclusivamente de la secuencia temporal mejora la robustez obtenida en clasificación de ataques variados.
- **Uso de técnicas para manejo del desbalance de clases (SMOTE, reweighting, ensambles):** La mayoría de los estudios aborda el desbalance mediante oversampling, pesos de clase o ensambles ponderados. Estas técnicas ayudan, pero no modifican la arquitectura del modelo,

por lo que el aprendizaje sigue siendo global y poco especializado a pesar de que ha habido aportes interesantes en esta área como investigaciones de métodos basados en SMOTE y GMMs (Gaussian Mixture Models) o en el uso de ensambles basados en soft-voting [4][5]. **Aporte del proyecto:** El presente proyecto busca generar un aporte del lado del modelo mediante la propuesta de una arquitectura de MoE buscando que distintos expertos aprendan patrones específicos de clases minoritarias sin depender únicamente de las técnicas para el manejo del desbalance de clases

- **Métodos no supervisados y uso de encoders para detección de intrusiones** Enfoques basados en autoencoders, contrastive learning y modelos one-class eliminan la dependencia de etiquetas, pero suelen presentar dificultades para discriminar tipos específicos de ataques y menor interpretabilidad operativa ya que se generan grupos de ataques reales y tráfico benigno lo cual no da claridad real sobre la clasificación de cada posible ataque [6]. **Aporte del proyecto:** Se busca un enfoque supervisado robusto, compatible con una clasificación lo más exacta posible no tan solo de un ataque sino del tipo de ataque involucrado a través del presente proyecto llegando a proponer un modelo el cual se busca validar que pueda emplearse para dar tratamiento específico por cada tipo de ataque identificado.
- **Mixture of Experts en aplicaciones fuera del área de la ciberseguridad:** El enfoque Mixture of Experts ha sido ampliamente estudiado y validado en dominios como visión por computador, procesamiento de lenguaje natural y aprendizaje a gran escala, donde su principal ventaja radica en el cómputo condicional y la especialización de submodelos. En estos contextos, el mixture of expers permite escalar la capacidad del modelo sin incrementar proporcionalmente el costo computacional, activando únicamente un subconjunto de expertos relevantes para cada entrada. Modelos como Sparsely-Gated MoE basados en arquitecturas jerárquicas de expertos han demostrado mejoras sustanciales en precisión, eficiencia y capacidad de generalización [7]**.** **Aporte del proyecto:** Este trabajo busca trasladar y adaptar los beneficios comprobados del Mixture of Experts en otras áreas de la IA hacia el dominio de la ciberseguridad, proponiendo una arquitectura a validar diseñada específicamente para identificar intrusiones en ataques de red a manera de dar valor a la teoría de este tipo de modelo dentro de la ciberseguridad.
- **Primeros trabajos con Mixture of Experts en tráfico 5G real** El estudio de Ilias et al. representa uno de los primeros esfuerzos formales en integrar una arquitectura Mixture of Experts dispersa posterior a una CNN para la detección de intrusiones en redes 5G reales. En su propuesta, una red convolucional profunda es utilizada como extractor de características, transformando el tráfico de red en representaciones las cuales son posteriormente procesadas por un Mixture of Expers con enrutamiento dinámico, donde solo

un subconjunto reducido de expertos es activado por muestra. Los experimentos se realizaron sobre datasets reales y representativos del contexto 5G, como 5G-NIDD y NANCY, demostrando que la arquitectura CNN + Mixture of Experts supera a modelos CNN tradicionales y a clasificadores profundos estándar tanto en métricas globales como en eficiencia computacional. Además, el uso de expertos especializados permitió una mejor discriminación entre distintos tipos de tráfico legítimo y malicioso. No obstante, el propio estudio identifica limitaciones importantes. En particular, el rendimiento disminuye notablemente en ataques de bajo volumen y en clases con muy baja representación, ya que el mecanismo de enrutamiento tiende a favorecer expertos entrenados con clases mayoritarias [8].
Aporte del proyecto: El proyecto extiende directamente la línea de investigación iniciada por Ilias et al., a través de la propuesta y validación de una nueva arquitectura basada en Mixture of Experts explícitamente orientada al manejo del desbalance extremo, con el objetivo de mejorar la detección de ataques raros y de bajo volumen favoreciendo la generalización.

5. Metodología de Trabajo

La metodología de trabajo propuesta para el desarrollo del proyecto se fundamenta en un enfoque iterativo, estructurado y orientado tanto a la investigación como a la experimentación, combinando de manera paralela actividades de estudio teórico con tareas prácticas de implementación y validación. Este enfoque busca garantizar que cada decisión técnica esté respaldada por fundamentos teóricos sólidos y por prácticas validadas en trabajos previos del estado del arte.

En una etapa inicial, y de forma transversal a todo el proyecto, se llevará a cabo una revisión exhaustiva de literatura científica, documentación técnica y artículos especializados relacionados con detección de intrusiones, aprendizaje automático en escenarios de clases desbalanceadas y arquitecturas avanzadas como Mixture of Experts. Esta revisión permitirá adquirir los conocimientos necesarios del área de estudio, identificar enfoques actuales, metodologías empleadas y métricas relevantes, así como establecer una base conceptual sólida para el desarrollo del proyecto.

De manera operativa, por cada jornada de trabajo se destinará un bloque inicial de entre dos y tres horas a la lectura y análisis de información relevante, previo a la ejecución de actividades de desarrollo. Posteriormente, se dedicará el tiempo restante a la implementación práctica del proyecto. Si bien durante el desarrollo podrán surgir dudas puntuales que requieran investigación adicional, estas corresponderán a aspectos específicos, dado que previamente se habrá realizado una investigación profunda alineada con cada una de las fases del proyecto. Por último, al ser el presente un proyecto de

carácter computacional se ira constantemente cargando los avances en un repositorio de GitHub a manera de ir registrando el progreso.

Asimismo, se considerarán los principios del diseño de ingeniería, tales como modularidad, escalabilidad, reproducibilidad y mantenibilidad, así como las buenas prácticas relacionadas con proyectos de machine learning, data engineering y desarrollo de software, asegurando la calidad técnica y la validez de los resultados obtenidos.

Para una mejor organización y control del proyecto, la metodología se divide en las siguientes fases:

Fase 1: Data Engineering

Esta fase tiene como objetivo principal preparar y estructurar adecuadamente los datos que servirán como insumo para el entrenamiento y evaluación de los modelos. Las actividades se ejecutarán de forma secuencial, dado que cada etapa depende del resultado de la anterior. Las actividades contempladas son:

- Selección del dataset de detección de intrusiones.
- Análisis estadístico del desbalance de clases y caracterización del dataset.
- Limpieza de datos.
- Preprocesamiento de datos.
- Selección de características (Feature Selection).
- Partición del dataset en conjuntos de entrenamiento, validación y prueba.
- Presentación del primer entregable.

Durante esta fase, el foco central será el manejo y preparación de los datos, por lo que la investigación se orientará a analizar cómo proyectos similares han abordado el preprocesamiento de datasets altamente desbalanceados. Para ello, se utilizarán como referencia las documentaciones oficiales y buenas prácticas asociadas a librerías de Python especializadas en procesamiento de datos, tales como Pandas y NumPy. El dataset final será descargado en un entorno local para respaldo y posteriormente cargado en un servicio como Lightning AI o Google Colab desde donde se pueda llevar a cabo la generación de los modelos de inteligencia artificial. Siguiendo estas actividades y alineándose con estándares utilizados en proyectos de detección de intrusiones, se garantizará un manejo adecuado y consistente de los datos.

Fase 2: Modelos Base y Técnicas Iniciales para Mitigar el Desbalanceo

El objetivo de esta fase es establecer una línea base de desempeño mediante la implementación de modelos tradicionales, así como evaluar técnicas clásicas para el manejo del desbalance de clases. Las actividades se desarrollarán de manera secuencial y comprenden:

- Implementación de modelos base no basados en Mixture of Experts.
- Entrenamiento y evaluación inicial de los modelos baseline.
- Aplicación de técnicas de reweighting de clases.
- Implementación y evaluación de Focal Loss en escenarios de clasificación multiclase.
- Aplicación de técnicas de oversampling, como SMOTE.
- Comparación del desempeño entre modelos base y técnicas de mitigación del desbalance.
- Presentación del segundo entregable.

Esta fase permitirá identificar las limitaciones de los enfoques tradicionales frente al desbalance extremo y establecer métricas de referencia que serán utilizadas posteriormente para comparar el desempeño del modelo Mixture of Experts. Para esta sección será necesario por un lado recurrir a la bibliografía teórica de Machine Learning para validar el uso de los modelos y técnicas de manejo de desbalanceo y por otro lado será necesaria la consulta continua de la librería de Machine Learning Scikit-Learn y PyTorch para verificar mediante que funciones de dichas librerías se pueden implementar los modelos base.

Fase 3: Diseño e Implementación de la Arquitectura Mixture of Experts

En esta fase se abordará el núcleo del proyecto, centrado en el diseño e implementación de una arquitectura Mixture of Experts adaptada al problema de detección de intrusiones con clases altamente desbalanceadas. Las actividades consideradas son:

- Diseño conceptual de la arquitectura Mixture of Experts.
- Implementación inicial del conjunto de expertos especializados.
- Implementación del mecanismo de gating.
- Definición e implementación de mecanismos de enrutamiento de instancias.
- Implementación de mecanismos de regularización para el balance de carga entre expertos.
- Definición e implementación de estrategias de especialización de expertos.
- Integración de Focal Loss dentro de la arquitectura Mixture of Experts.
- Entrenamiento y validación del modelo.
- Presentación del tercer entregable

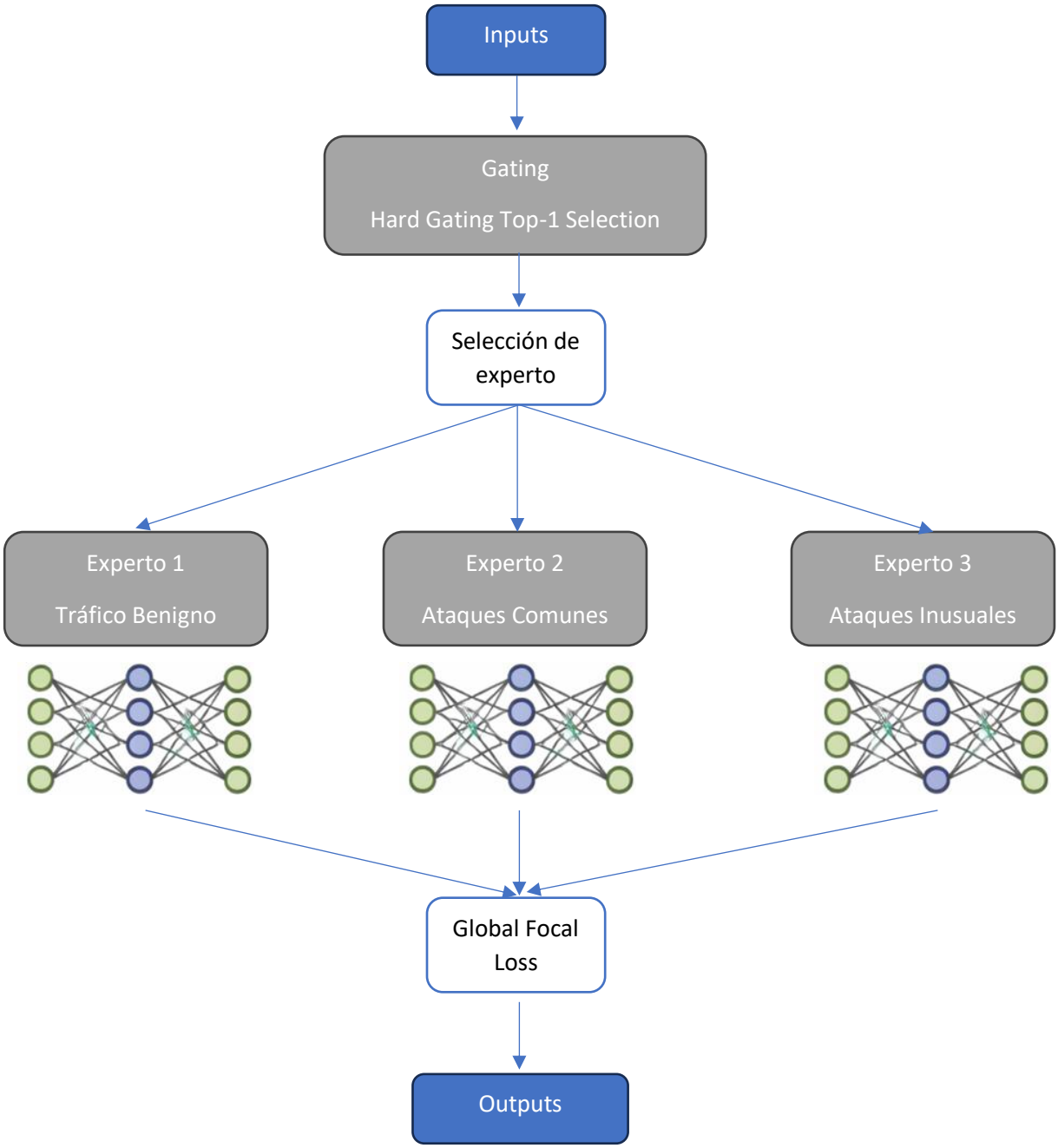
Durante esta fase se priorizará la especialización de los expertos y el análisis del comportamiento del mecanismo de enrutamiento, asegurando que el modelo aproveche adecuadamente la heterogeneidad de los datos y mejore el desempeño en clases minoritarias. En esta sección será la que más dedicación se invertirá a la investigación y se buscará mostrar avances constantes al tutor a manera de ir avanzando continuamente respecto a esta línea la cual es el eje central del proyecto

Con base en lo investigado, se propone para el presente proyecto una arquitectura Mixture of Experts de carácter experimental, cuyo diseño inicial integra hard-gating, Focal Loss, expertos homogéneos y un esquema de entrenamiento por fases. Si bien la arquitectura se plantea de forma coherente con base en el estado del arte, su desempeño no se asume como garantizado. Por ello, durante el desarrollo del proyecto se priorizarán la estabilidad del entrenamiento, la evaluación rigurosa de resultados y el análisis del comportamiento del modelo, permitiendo realizar simplificaciones o ajustes controlados en la arquitectura en caso de ser necesario, con el fin de validar de manera objetiva si las decisiones de diseño propuestas contribuyen efectivamente a mejorar la clasificación de intrusiones en escenarios de desbalance extremo. El detalle de dicha arquitectura es presentado a continuación:

- **Arquitectura Hard-Gating con top-k=1:** Se propone emplear un esquema hard gating con top-k=1 a evaluar bajo la hipótesis de que restringir la contribución a un único experto altamente especializado podría favorecer una clasificación más precisa para ciertos tipos de ataques inusuales.
- **Cantidad de expertos= 3 o =4 con base en tests**: Dado el fuerte desbalance del dataset, un número elevado de expertos podría afectar la calidad del entrenamiento, ya que algunas clases no cuentan con suficientes datos. Por ello, se propone una arquitectura con 3 expertos: uno para tráfico benigno, otro para ataques comunes (p. ej., DDoS) y otro para ataques menos frecuentes (p. ej., Botnet). Si esta configuración limita la granularidad en la clasificación, se evaluará una alternativa con 4 expertos, manteniendo uno para tráfico benigno y tres especializados por familias de ataques. La validación de la arquitectura experimental permitirá determinar si la cantidad de expertos seleccionada contribuye efectivamente a mejorar la clasificación.
- **Focal Loss aplicado sobre la salida final:** Se plantea como hipótesis a validar que el mecanismo de gating debe volverse más especializado en la selección del experto adecuado, particularmente para el tratamiento de clases minoritarias. Con este objetivo, se propone aplicar Focal Loss en la salida final del modelo, con el fin de penalizar menos las clases mayoritarias y enfatizar el aprendizaje sobre ejemplos difíciles o poco representados, buscando demostrar si esto favorece a una mejor discriminación de ataques minoritarios.

- **Expertos homogéneos entrenados con datos heterogéneos:** Se propone emplear expertos homogéneos bajo la hipótesis de que la especialización ocurra principalmente por el contenido de los datos y no por diferencias en la capacidad de cómputo de los modelos. Esto se complementará con que los expertos serán entrenados con datos heterogéneos mediante subsets de datos basados en la siguiente organización: dataset original para entrenar experto de tráfico benigno y datasets rebalanceados con SMOTE para entrenar expertos dependiendo el tipo de tráfico en el que se busca especialización. Al final se validará con una matriz de utilización y comparación con modelos baselines si esta propesta beneficia la especialización.
- **Modelo para los expertos- red neuronal MLP profunda:** En reconocimiento de que los datos de tráfico son tabulares y que por ende el objetivo sería buscar correlaciones entre las features se sugiere que los expertos sean de tipo red neuronal MLP con capas densas totalmente conectadas para modelar combinaciones no lineales entre las características del tráfico de red y se propone el uso de la función de activación LeakyRELU para evitar que los gradientes de las clases minoritarias desaparezcan. Esta propuesta contrasta con investigaciones de propósitos similares de clasificación de intrusiones en donde se ha empleado CNNs de 1D y 2D otorgando resultados aceptables, pero dentro del presente proyecto se buscará validar si el uso de redes MLP profundas mejora o no las métricas objetivo.
- **Entrenamiento de tres fases:** El entrenamiento del modelo se plantea en tres fases. En una primera etapa, cada experto se entrena de forma independiente para favorecer su especialización en tipos específicos de ataques. En una segunda fase, se congelan los expertos y se entrena únicamente el mecanismo de gating utilizando Focal Loss global, con el objetivo de mejorar la selección del experto adecuado, especialmente para clases minoritarias. Finalmente, se realiza una fase de refinamiento conjunto, en la que se descongelan los expertos y se entrena toda la arquitectura Mixture of Experts con una tasa de aprendizaje mayor. Esta estrategia se propone bajo la hipótesis de que mejora tanto la especialización de los expertos como la capacidad del gating, hipótesis que será evaluada durante la validación del modelo.
- **Consideración a la especialización:** Para considerar la especialización se plantea el diseño de una matriz de utilización de expertos por clase en la cual para cada clase se indica en porcentaje las muestras de dicha clase que fueron asignadas por el gating a dicho experto de modo que se reconozca si el gating está asignado correctamente a cada experto. Y a su vez se buscará determinar la especialización de los expertos para lo cual se considerará su recall por clase en comparación a los modelos baseline para probar si hay una ganancia respecto al modelo base.

Por último, se resume toda la arquitectura anteriormente detallada mediante la siguiente imagen:



Fase 4: Evaluación y Análisis de Modelos

El objetivo de esta fase es realizar una evaluación exhaustiva y comparativa del modelo propuesto. Las actividades incluyen:

- Evaluación del modelo Mixture of Experts utilizando métricas apropiadas para clases desbalanceadas.
- Comparación del desempeño del modelo Mixture of Experts frente a los modelos baseline.

- Análisis del proceso de enrutamiento del modelo final.
- Análisis de la distribución de carga y utilización de cada experto.

Este análisis permitirá interpretar el comportamiento interno del modelo, validar las hipótesis planteadas inicialmente y justificar los beneficios del enfoque propuesto frente a métodos tradicionales. En esta parte, el foco será revisar como en investigaciones o proyectos similares se ha desarrollado la valoración de los modelos con base en métricas de interés (p.e. F1-Score y Recall) a manera de tomar como bases dichas comparativas para hacer un análisis que sea sustentado en datos y gráficos que puedan evidenciar claramente la calidad de los modelos.

El objetivo central de la validación es demostrar la superioridad de la arquitectura Mixture of Experts frente a clasificadores baseline en un entorno de desbalance extremo. La validación se articulará mediante dos partes: métricas predictivas y análisis de especialización. Cabe destacar, que estas validaciones permitirán comprobar o no si las hipótesis del diseño de la arquitectura benefician o no a la clasificación de intrusiones en un dataset desbalanceado:

Para la parte de la validación mediante métricas se usarán las siguientes:

- **Recall por Clase:** El recall nos permite disminuir los falsos negativos ya que estos serían los casos más críticos en los cuales diríamos que un tráfico no es un ataque cuando en realidad si lo es dejando que el mismo pase
- **F1-Score Macro:** Esta métrica es importante ya que nos permite ponderar la métrica F1 por cada clase pudiendo evaluar el modelo por clases dando igual importancia a clases minoritarias como a clases mayoritarias

A su vez, a manera de complemento se pueden llevar a cabo matrices de confusión y curvas precisión-recall para ver fronteras de decisión y con eso poder comparar de manera correcta los modelos base contra el MoE pudiendo de esa manera medir la ganancia del modelo de expertos pudiendo indicar si un MoE con Focal Loss mejora a un modelo baseline con ajuste adecuado.

Por otro lado, para el análisis de especialización de igual forma se buscará hacer la tarea de comparación de MoE contra baselines ya que como se ha indicado el mismo se estructurará de la siguiente forma: generación de una matriz de utilización de expertos, la cual cuantificará el porcentaje de asignación de cada clase a un experto específico por parte del gating y complementariamente, se evaluará la ganancia comparando el recall por clase del MoE frente a los modelos baseline donde el éxito quedará

determinado por un incremento significativo en esta métrica, demostrando que el experto ha desarrollado una capacidad de detección superior a la del modelo base.

Al finalizar el proceso de evaluación, el proyecto se considerará exitoso si la arquitectura Mixture of Experts, ya sea en su diseño inicial o tras ajustes derivados de los experimentos y validaciones realizadas, demuestra una mejora consistente y medible en las métricas objetivo de recall por clase y F1-score macro en comparación con los modelos baseline configurados mediante un proceso controlado de selección de hiperparámetros. Dicha mejora deberá alcanzarse sin una degradación significativa del desempeño global del modelo, y evidenciando además una mejor capacidad de generalización, particularmente al favorecer la correcta clasificación de clases minoritarias.

Fase 5: Preparaciones Finales y Conclusión del Proyecto

La fase final estará orientada al cierre técnico y documental del proyecto, así como a la presentación de los resultados obtenidos. Las actividades contempladas son:

- Redacción del documento final del proyecto.
- Preparación y presentación final del proyecto.
- Organización del repositorio y de los entregables finales.

Con esta fase se busca documentar todo el proceso seguido siguiendo la normativa IEEE para el formato adecuado del documento y complementando el mismo con el formato requerido por la biblioteca la USFQ. Finalmente se llevará a cabo la presentación del proyecto en la cual se sintetizará todo lo desarrollado buscando centrarse en los resultados, validaciones, aprendizajes y aportes respecto al tema.

6. Sumario de Contenidos

- Abstract
- Introducción
- Estado del arte
- Descripción de la Propuesta
- Desarrollo del Proyecto: Parte 1. Data Engineering
- Desarrollo del Proyecto: Parte 2. Modelos base y técnicas iniciales para mitigar desbalanceo
- Desarrollo del Proyecto: Parte 3. Implementación y Arquitectura de Mixture of Experts
- Análisis de resultados
- Conclusiones y Recomendaciones
- Trabajo futuro

7. Recursos

a. Humanos

Estudiante (John Ochoa Abad) y tutor (Ricardo Flores Moyano)

b. Materiales

Hardware: Laptop Personal HP Modelo ZBook Firefly 14, Computador Personal Customizado con 16GB de RAM, Intel Core I7 9700K y GPU NVIDIA RTX 2070 Super, High Performance Computing Center USFQ

Software: Lenguajes: Python, SQL

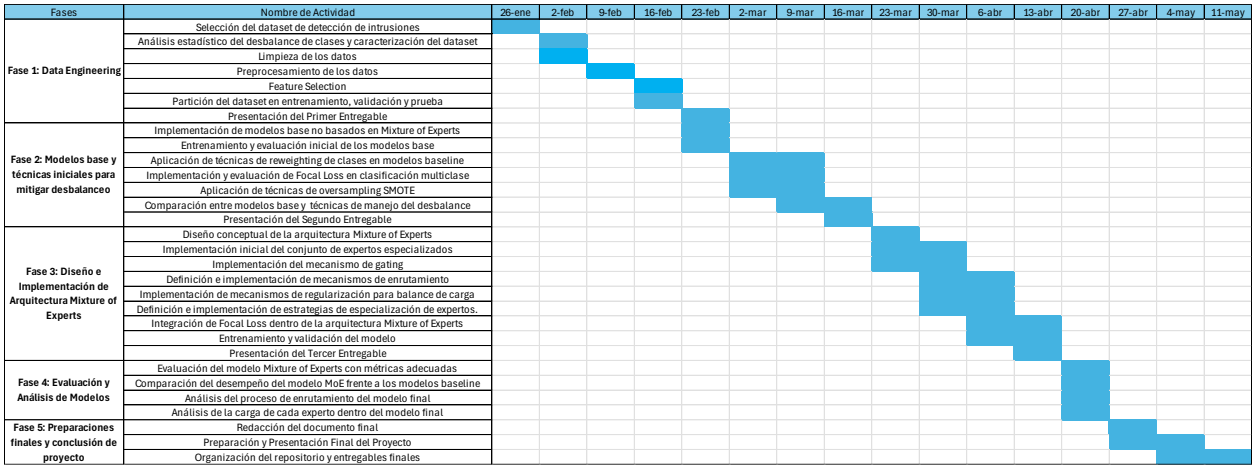
IDE / Entorno: VS Code**,** Google Colab**,** Lightning AI

Librerías / Frameworks: NumPy**,** Pandas**,** Scikit-learn**,** PyTorch

c. Económicos

El entorno de Lightning IA tiene costo de GPU por hora con un valor de $0.53 por hora de uso de un GPU de modelo NVIDIA T4 con el entrenamiento del modelo se pueden llegar a requerir alrededor de 10 iteraciones de entrenamientos y por experiencia algún entrenamiento puede llegar a tomar hasta 12 horas por lo que se requeriría un valor de $63.6 para cubrir los costos.

8. Cronograma de Actividades



9. Entregables

Como entregables del proyecto se plantea lo siguiente:

- Export de los datos preparados y usados en el entrenamiento del modelo mixture of experts
- Repositorio Github con el código final empleado en todo el proyecto
- Export del modelo final entrenado
- Tablas y Gráficos de Métricas del modelo final
- Documento formal del proyecto desarrollado siguiendo lineamientos de IEEE

10. Referencias

[1] A. Thakkar and R. Lohiya, "Fusion of statistical importance for feature selection in deep neural network-based intrusion detection system," *Information Fusion*, vol. 90, pp. 353–363, 2023, doi: 10.1016/j.inffus.2022.09.026.

[2] U. K. Lilhore, S. Dalal, and S. Simaiya, "A cognitive security framework for detecting intrusions in IoT and 5G utilizing deep learning," *Computers & Security*, vol. 136, p. 103560, 2024, doi: 10.1016/j.cose.2023.103560.

[3] T. E. T. Djaidja, B. Brik, S. M. Senouci, A. Boualouache, and Y. Ghamri-Doudane, "Early network intrusion detection enabled by attention mechanisms and RNNs," IEEE Transactions on Information Forensics and Security, vol. 19, pp. 7783–7793, 2024, doi: 10.1109/TIFS.2024.3441862.

[4] H. Zhang, L. Huang, C. Q. Wu, and Z. Li, "An effective convolutional neural network based on SMOTE and Gaussian mixture model for intrusion detection in imbalanced dataset," *Computer Networks*, vol. 177, p. 107315, 2020, doi: 10.1016/j.comnet.2020.107315.

[5] R. Almuhanna and S. Dardouri, "A deep learning/machine learning approach for anomaly based network intrusion detection," *Frontiers in Artificial Intelligence*, vol. 8, p. 1625891, 2025, doi: 10.3389/frai.2025.1625891.

[6] L. Yuan *et al.*, "Manticore: An unsupervised intrusion detection system based on contrastive learning in 5G networks," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Seoul, South Korea, 2024, pp. 4705–4709, doi: 10.1109/ICASSP48485.2024.10447814.

[7] Y. Han *et al.*, "Dynamic neural networks: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 11, pp. 7436–7456, 2022, doi: 10.1109/TPAMI.2021.3117837.

[8] L. Ilias, G. Doukas, V. Lamprou, C. Ntanos, and D. Askounis, "Convolutional neural networks and mixture of experts for intrusion detection in 5G networks and beyond," *Frontiers in Artificial Intelligence*, vol. 8, Art. no. 1708953, 2026, doi: 10.3389/frai.2025.1708953.

11. Revisión y firma del tutor del proyecto

Yo, Ricardo Flores Moyano, profesor de la carrera de Ingeniería en Ciencias de la Computación, hago constar que he revisado y, por lo tanto, apruebo el documento de planificación del proyecto titulado "Arquitectura Mixture of Experts para Clasificación de Ataques en Datasets con Desbalance Extremo de Clases" propuesto por el estudiante John Ochoa Abad. Por otra parte, me comprometo a proporcionar al estudiante el soporte necesario y oportuno para el buen desarrollo del proyecto antes mencionado.

_____

Fdo: Ricardo Flores Moyano

Quito, 30 de Enero de 2026