

Sistema automático de predicción de incidencias en redes de difusión de TV digital

Lorena Álvarez-Pérez, Francisco J. González-Serrano
Marcelino Lázaro, Harold Molina-Bulla

Marcelino Lázaro

Departamento de Teoría de la Señal y Comunicaciones

www.tsc.uc3m.es/~mlazaro/

Proyecto PID2021-125652OB-I00 (ML-SAM) financiado por



Índice de contenidos

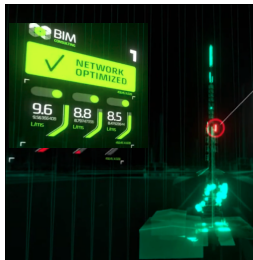
- Introducción
 - ▶ Alarmas e Incidencias
 - ▶ Sistema de difusión de TV digital
- Sistema de predicción automática de incidencias
- Modelos de aprendizaje empleados
- Resultados
- Conclusiones

Alarmas / Incidencias

Red de Difusión Terrestre



Estación Transmisora



Centro Operación (NOC)

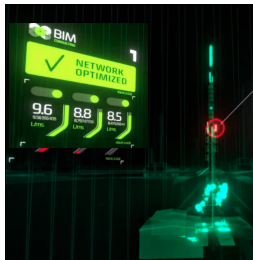


Alarmas / Incidencias

Red de Difusión Terrestre



Estación Transmisora



Centro Operación (NOC)



BD Alarmas

Alarma

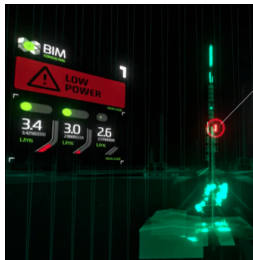


Alarmas / Incidencias

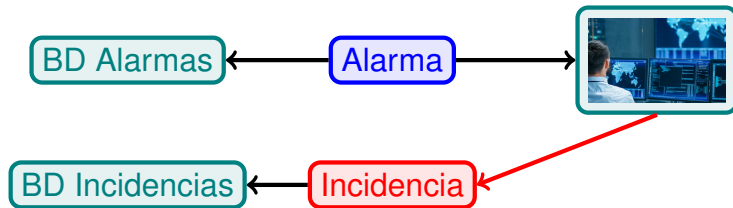
Red de Difusión Terrestre



Estación Transmisora



Centro Operación (NOC)

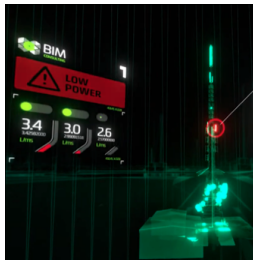


Alarmas / Incidencias

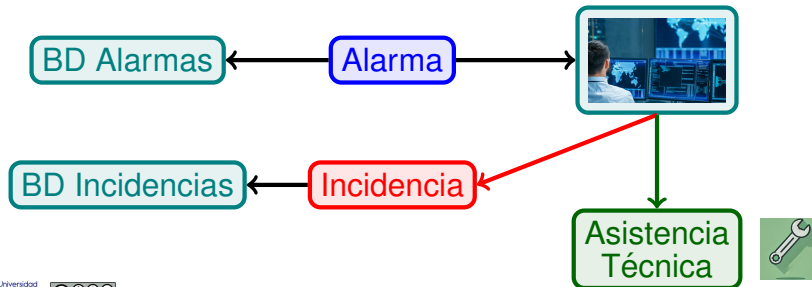
Red de Difusión Terrestre



Estación Transmisora



Centro Operación (NOC)

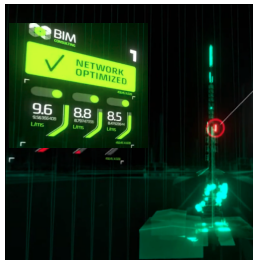


Alarmas / Incidencias

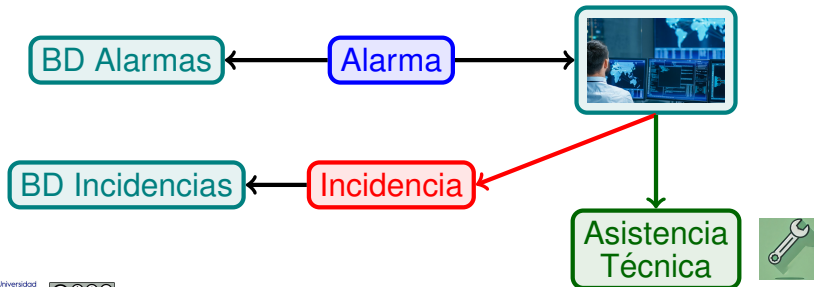
Red de Difusión Terrestre



Estación Transmisora



Centro Operación (NOC)

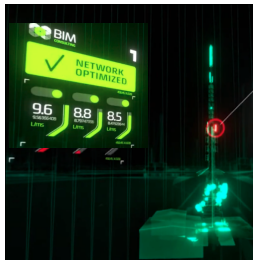


Alarmas / Incidencias

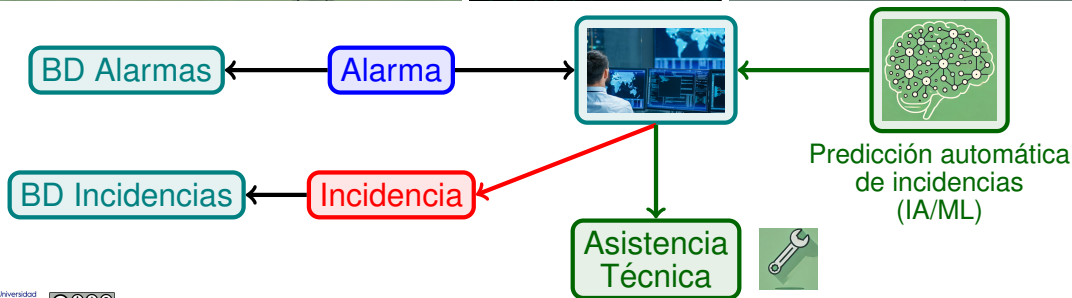
Red de Difusión Terrestre



Estación Transmisora



Centro Operación (NOC)



Sistema de difusión de TV digital

● Contribución

- ▶ Recogida de contenido de los productores
- ▶ Agrupación de programas en “servicios” (multiplex de varios programas)



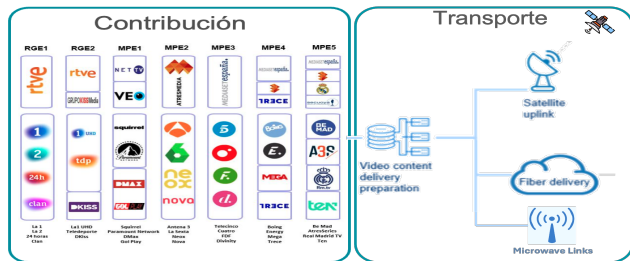
Sistema de difusión de TV digital

- Contribución

- ▶ Recogida de contenido de los productores
- ▶ Agrupación de programas en “servicios” (multiplex de varios programas)

- Transporte

- ▶ Se puede realizar por redes IP, enlaces de satélite o enlaces de microondas



Sistema de difusión de TV digital

- Contribución

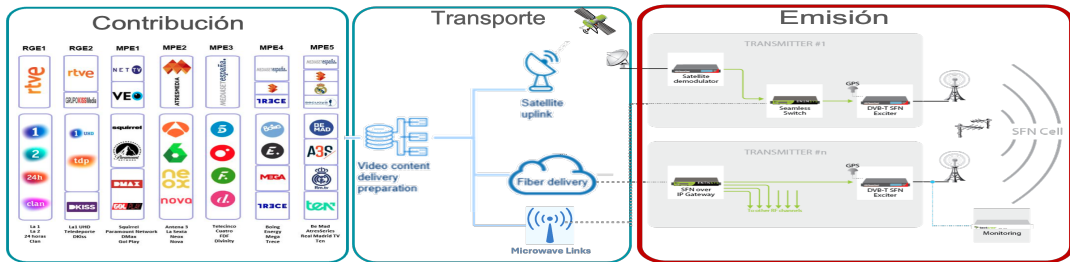
- ▶ Recogida de contenido de los productores
- ▶ Agrupación de programas en “servicios” (multiplex de varios programas)

- Transporte

- Se puede realizar por redes IP, enlaces de satélite o enlaces de microondas

- Emisión

- Se demultiplexan las contribuciones para extraer cada programa, seguido de modulación, amplificación RF y transmisión en el canal correspondiente

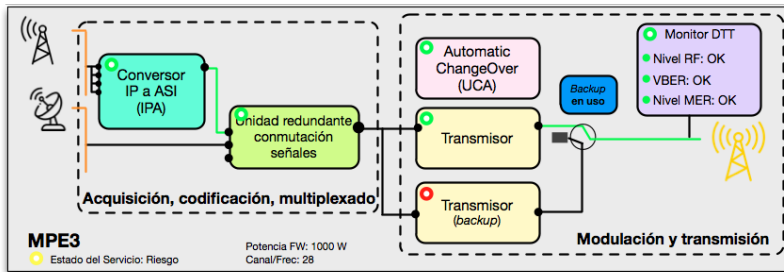


Emisión : Servicios y elementos en cada estación transmisora

- Servicios (Multiplex): RGE1, RGE2, MPE1, MPE2, MPE3...

Emisión : Servicios y elementos en cada estación transmisora

- Servicios (Multiplex): RGE1, RGE2, MPE1, MPE2, MPE3...

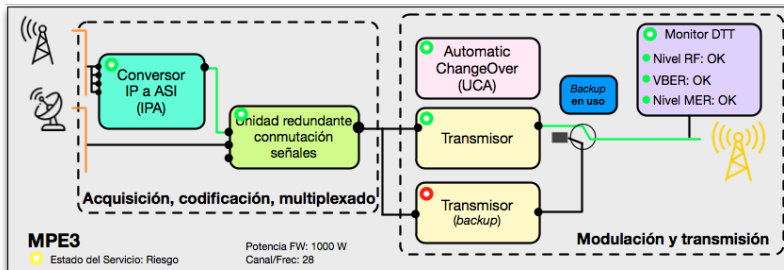


- Elementos por cada servicio (p.e. MPE3)

- ▶ Banda base
 - ★ IPA, Conmutador, Chasis/reloj/GPS
- ▶ RF
 - ★ N transmisores (1 por servicio) + 1 de backup
 - ★ UCA, DTT monitor
- ▶ Sistema de energía y climatización de la estación

Emisión : Servicios y elementos en cada estación transmisora

- Servicios (Multiplex): RGE1, RGE2, MPE1, MPE2, MPE3...



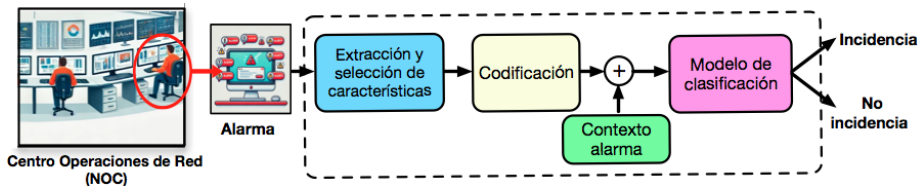
- Elementos por cada servicio (p.e. MPE3)

- ▶ Banda base
 - ★ IPA, Conmutador, Chasis/reloj/GPS
- ▶ RF
 - ★ N transmisores (1 por servicio) + 1 de backup
 - ★ UCA, DTT monitor
- ▶ Sistema de energía y climatización de la estación

- Alarmas : de elemento / de servicio

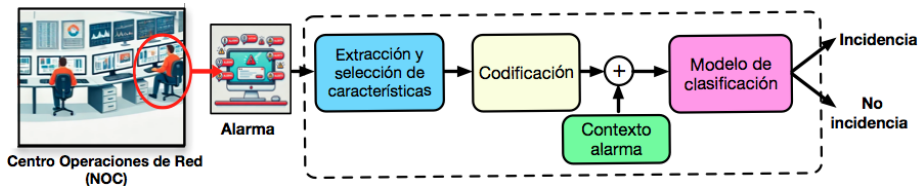
Sistema de predicción automática de incidencias

- Procesado de la alarma: formato común para elementos y servicios (57 campos)
 - ▶ Extracción y selección de características
 - ▶ Codificación según el tipo de campo
 - ★ Numérico, binario, categórico, texto, diccionario (info elementos y servicios, contratos, SLAs)



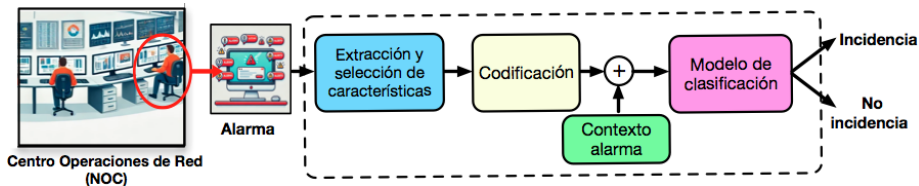
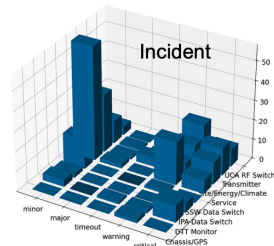
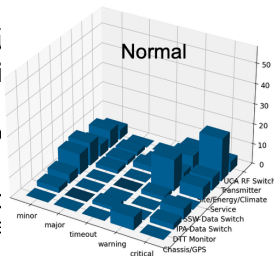
Sistema de predicción automática de incidencias

- Procesado de la alarma: formato común para elementos y servicios (57 campos)
 - ▶ Extracción y selección de características
 - ▶ Codificación según el tipo de campo
 - ★ Numérico, binario, categórico, texto, diccionario (info elementos y servicios, contratos, SLAs)
- Información contextual
 - ▶ Estadísticos temporales de alarmas por servicio y elemento
 - ★ Número, duración, min, max, media, mediana, desviación...



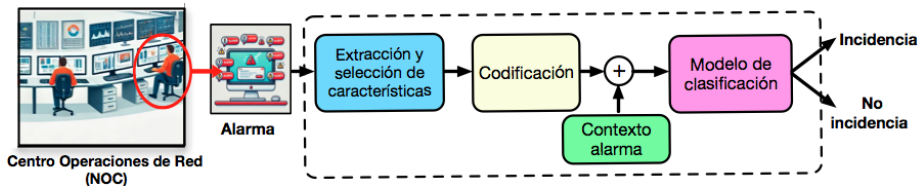
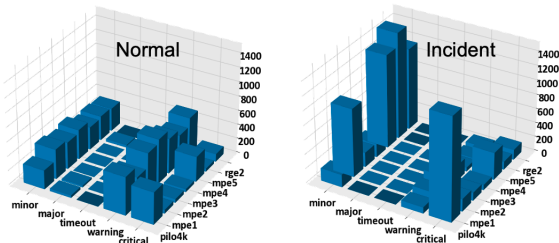
Sistema de predicción automática de incidencias

- Procesado de la alarma: formato común
 - ▶ Extracción y selección de características
 - ▶ Codificación según el tipo de campo
 - ★ Numérico, binario, categórico, texto, diccionario
- Información contextual
 - ▶ Estadísticos temporales de alarmas p...
 - ★ Número, duración, min, max, media, me...



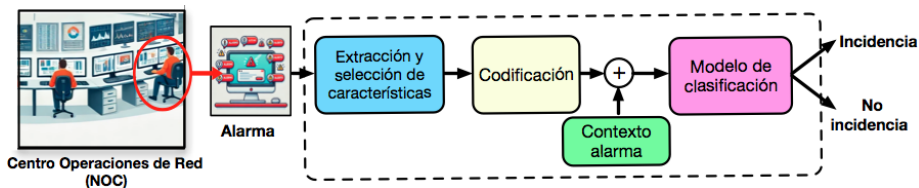
Sistema de predicción automática de incidencias

- Procesado de la alarma: formato común
 - ▶ Extracción y selección de características
 - ▶ Codificación según el tipo de campo
 - ★ Numérico, binario, categórico, texto, dic
- Información contextual
 - ▶ Estadísticos temporales de alarmas
 - ★ Número, duración, min, max, media, me



Sistema de predicción automática de incidencias

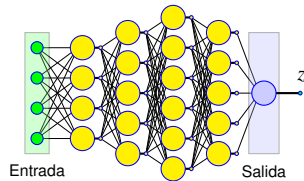
- Procesado de la alarma: formato común para elementos y servicios (57 campos)
 - ▶ Extracción y selección de características
 - ▶ Codificación según el tipo de campo
 - ★ Numérico, binario, categórico, texto, diccionario (info elementos y servicios, contratos, SLAs)
- Información contextual
 - ▶ Estadísticos temporales de alarmas por servicio y elemento
 - ★ Número, duración, min, max, media, mediana, desviación...
 - ▶ Patrones en la secuencia de alarmas
 - ★ Análisis de secuencias frecuentes (PrefixSpan top- k patterns)
 - ★ Métricas de alineamiento: similaridad con patrones de incidencia o no incidencia



Modelos de aprendizaje supervisado

- Solución propuesta en el proyecto (clasificador binario)
 - ▶ MLP con función de coste Bayesiana

$$L^{Bayes}(\mathbf{w}) = \bar{c}_0 \hat{p}_{FA} + \bar{c}_1 \hat{p}_M$$



Modelos de aprendizaje supervisado

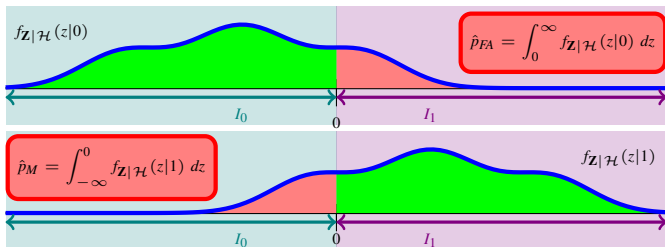
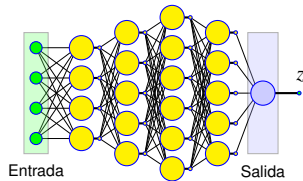
- Solución propuesta en el proyecto (clasificador binario)

- MLP con función de coste Bayesiana

$$L^{Bayes}(\mathbf{w}) = \bar{c}_0 \hat{p}_{FA} + \bar{c}_1 \hat{p}_M$$

- ★ Estimaciones de las probabilidades de falsa alarma y de pérdida

$$\hat{p}_{FA} = \int_0^{\infty} \hat{f}_{\mathbf{Z}|\mathcal{H}}(z|0) dz \quad \hat{p}_M = \int_{-\infty}^0 \hat{f}_{\mathbf{Z}|\mathcal{H}}(z|1) dz$$

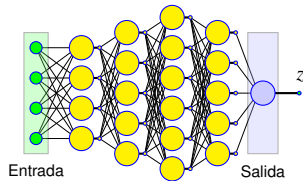


Modelos de aprendizaje supervisado

- Solución propuesta en el proyecto (clasificador binario)

- ▶ MLP con función de coste Bayesiana

$$L^{Bayes}(\mathbf{w}) = \bar{c}_0 \hat{p}_{FA} + \bar{c}_1 \hat{p}_M$$

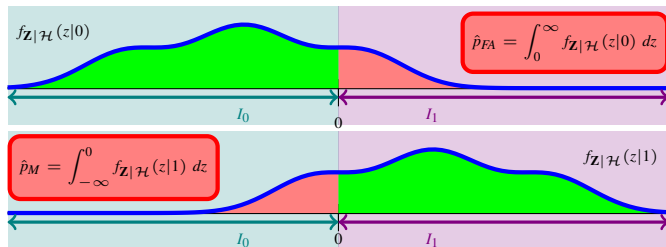
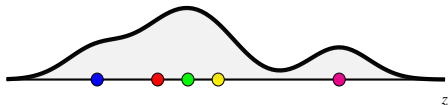
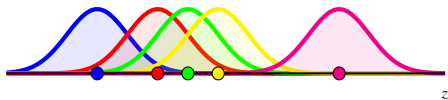


- ★ Estimaciones de las probabilidades de falsa alarma y de pérdida

$$\hat{p}_{FA} = \int_0^\infty \hat{f}_{\mathbf{Z}|\mathcal{H}}(z|0) dz \quad \hat{p}_M = \int_{-\infty}^0 \hat{f}_{\mathbf{Z}|\mathcal{H}}(z|1) dz$$

- ★ Estimador de Parzen: distribuciones condicionales de la salida de la red

$$\hat{f}_{\mathbf{Z}|\mathcal{H}}(z|t) = \frac{1}{N_t} \sum_{k \in S_t} k_\sigma(z - z_k), \quad t \in \{0, 1\}$$



Modelos de aprendizaje supervisado

- Solución propuesta en el proyecto (clasificador binario)

- ▶ MLP con función de coste Bayesiana

$$L^{Bayes}(\mathbf{w}) = \bar{c}_0 \hat{p}_{FA} + \bar{c}_1 \hat{p}_M$$

- ★ Estimaciones de las probabilidades de falsa alarma y de pérdida

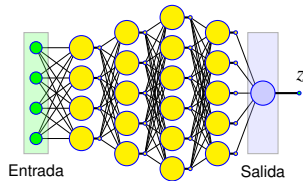
$$\hat{p}_{FA} = \int_0^\infty \hat{f}_{\mathbf{Z}|\mathcal{H}}(z|0) dz \quad \hat{p}_M = \int_{-\infty}^0 \hat{f}_{\mathbf{Z}|\mathcal{H}}(z|1) dz$$

- ★ Estimador de Parzen: distribuciones condicionales de la salida de la red

$$\hat{f}_{\mathbf{Z}|\mathcal{H}}(z|t) = \frac{1}{N_t} \sum_{k \in S_t} k_\sigma(z - z_k), \quad t \in \{0, 1\}$$

- Otros métodos evaluados

- ▶ Logistic Regression (LR)
- ▶ Asymmetric Label Switched Ensemble (ALSE)
- ▶ LightGBM
- ▶ RUSBoost



Simulador de alarmas

- El sistema se ha desarrollado en el ámbito del proyecto Smart-NOC (CDTI)
 - ▶ <https://www.cellnex.com/es-es/casos-uso/smart-noc/>
- También se ha realizado un simulador de alarmas
 - ▶ Simulación de diferentes escenarios (datos sintéticos)
 - ★ Baseline
 - ★ Degraded
 - ★ Alarm-Storm

	Información contexto	Características	Muestras		IR
			C_0	C_1	
Baseline	Sin contexto	540	9755	179	54,49
	PatSeqs	541			
	TimeStats	612			
	Pat+Time	613			
Degraded	Sin contexto	538	7049	1054	6,68
	PatSeqs	540			
	TimeStats	646			
	Pat+Time	648			
Alarm-storm	Sin contexto	538	3525	1499	2,35
	PatSeqs	540			
	TimeStats	610			
	Pat+Time	612			

Resultados

● Análisis: compromiso p_{FA} y p_D

- Mismo peso a los errores en ambas clases (Acierto Bal.)

	Ctx	Tipo		Acierto Bal.	P_{FA}	P_D
		Pat	Time			
LR	No	✓	X	0,9659	0,0261	0,9579
		✓	X	0,9682	0,0280	0,9646
	Yes	X	✓	0,9636	0,0320	0,9593
		✓	✓	0,9673	0,0292	0,9639
MLP Bayes	No	✓	X	0,9672	0,0229	0,9572
		✓	X	0,9677	0,0252	0,9608
	Yes	X	✓	0,9766	0,0236	0,9769
		✓	✓	0,9780	0,0228	0,9790
ALSE	No	✓	X	0,9670	0,0247	0,9586
		✓	X	0,9691	0,0264	0,9646
	Yes	X	✓	0,9752	0,0343	0,9847
		✓	✓	0,9783	0,0301	0,9867
Light GBM	No	✓	X	0,9637	0,0355	0,9620
		✓	X	0,9689	0,0349	0,9726
	Yes	X	✓	0,9860	0,0133	0,9853
		✓	✓	0,9888	0,0116	0,9893
RUS Boost	No	✓	X	0,9579	0,0390	0,9549
		✓	X	0,9621	0,0348	0,9590
	Yes	X	✓	0,9786	0,0173	0,9745
		✓	✓	0,9795	0,0167	0,9757

Alarm Storm

Resultados

- Análisis: compromiso p_{FA} y p_D
 - ▶ Mismo peso a los errores en ambas clases (Acierto Bal.)
- Sin contexto
 - ▶ MLP Bayes obtiene las mejores prestaciones

	Ctx	Tipo		Acierto Bal.	P_{FA}	P_D
		Pat	Time			
LR	No	-	-	0,9659	0,0261	0,9579
		✓	X	0,9682	0,0280	0,9646
	Yes	X	✓	0,9636	0,0320	0,9593
		✓	✓	0,9673	0,0292	0,9639
MLP Bayes	No	-	-	0,9672	0,0229	0,9572
		✓	X	0,9677	0,0252	0,9608
	Yes	X	✓	0,9766	0,0236	0,9769
		✓	✓	0,9780	0,0228	0,9790
ALSE	No	-	-	0,9670	0,0247	0,9586
		✓	X	0,9691	0,0264	0,9646
	Yes	X	✓	0,9752	0,0343	0,9847
		✓	✓	0,9783	0,0301	0,9867
Light GBM	No	-	-	0,9637	0,0355	0,9620
		✓	X	0,9689	0,0349	0,9726
	Yes	X	✓	0,9860	0,0133	0,9853
		✓	✓	0,9888	0,0116	0,9893
RUS Boost	No	-	-	0,9579	0,0390	0,9549
		✓	X	0,9621	0,0348	0,9590
	Yes	X	✓	0,9786	0,0173	0,9745
		✓	✓	0,9795	0,0167	0,9757

Alarm Storm

Resultados

- Análisis: compromiso p_{FA} y p_D
 - ▶ Mismo peso a los errores en ambas clases (Acierto Bal.)
- Sin contexto
 - ▶ MLP Bayes obtiene las mejores prestaciones
- La información contextual mejora la predicción
 - ▶ Los patrones de secuencias ofrecen mayor mejora que las estadísticas temporales
 - ★ Más acusado en datos sintéticos que en reales
 - ★ Combinación de ambas: mejora en prestaciones
 - ▶ Mejor método: LightGBM

	Ctx	Tipo		Acierto Bal.	P_{FA}	P_D
		Pat	Time			
LR	No	-	-	0,9659	0,0261	0,9579
	-	✓	X	0,9682	0,0280	0,9646
	Yes	X	✓	0,9636	0,0320	0,9593
	-	✓	✓	0,9673	0,0292	0,9639
MLP Bayes	No	-	-	0,9672	0,0229	0,9572
	-	✓	X	0,9677	0,0252	0,9608
	Yes	X	✓	0,9766	0,0236	0,9769
	-	✓	✓	0,9780	0,0228	0,9790
ALSE	No	-	-	0,9670	0,0247	0,9586
	-	✓	X	0,9691	0,0264	0,9646
	Yes	X	✓	0,9752	0,0343	0,9847
	-	✓	✓	0,9783	0,0301	0,9867
Light GBM	No	-	-	0,9637	0,0355	0,9620
	-	✓	X	0,9689	0,0349	0,9726
	Yes	X	✓	0,9860	0,0133	0,9853
	-	✓	✓	0,9888	0,0116	0,9893
RUS Boost	No	-	-	0,9579	0,0390	0,9549
	-	✓	X	0,9621	0,0348	0,9590
	Yes	X	✓	0,9786	0,0173	0,9745
	-	✓	✓	0,9795	0,0167	0,9757

Alarm Storm

Resultados

- Análisis: compromiso p_{FA} y p_D
 - ▶ Mismo peso a los errores en ambas clases (Acuerdo Bal.)
- Sin contexto
 - ▶ MLP Bayes obtiene las mejores prestaciones
- La información contextual mejora la predicción
 - ▶ Los patrones de secuencias ofrecen mayor mejora que las estadísticas temporales
 - ★ Más acusado en datos sintéticos que en reales
 - ★ Combinación de ambas: mejora en prestaciones
 - ▶ Mejor método: LightGBM
- Conclusión
 - ▶ La selección del modelo óptimo depende del punto de trabajo deseado por el operador (equilibrio entre p_{FA} y p_D)
 - ▶ MLP Bayes permite modificar el punto de trabajo de forma fundamentada
 - ★ Modificando los parámetros $\bar{c}_0$ y $\bar{c}_1$ en $L^{Bayes}(\mathbf{w})$

$$L^{Bayes}(\mathbf{w}) = \bar{c}_0 \hat{p}_{FA} + \bar{c}_1 \hat{p}_M$$

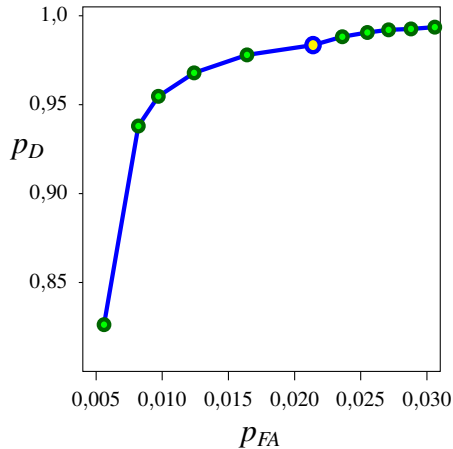
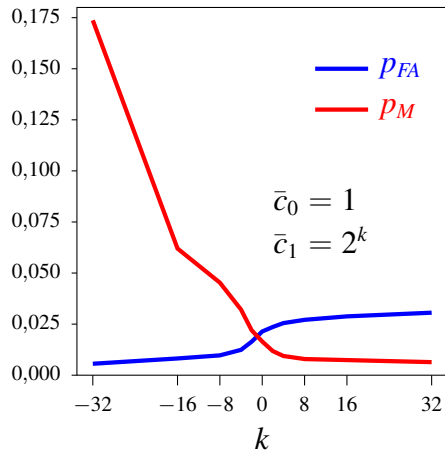
	Ctx	Tipo		Acuerdo Bal.	P_{FA}	P_D
		Pat	Time			
LR	No	-	-	0,9659	0,0261	0,9579
	-	✓	X	0,9682	0,0280	0,9646
	Yes	X	✓	0,9636	0,0320	0,9593
	-	✓	✓	0,9673	0,0292	0,9639
MLP Bayes	No	-	-	0,9672	0,0229	0,9572
	-	✓	X	0,9677	0,0252	0,9608
	Yes	X	✓	0,9766	0,0236	0,9769
	-	✓	✓	0,9780	0,0228	0,9790
ALSE	No	-	-	0,9670	0,0247	0,9586
	-	✓	X	0,9691	0,0264	0,9646
	Yes	X	✓	0,9752	0,0343	0,9847
	-	✓	✓	0,9783	0,0301	0,9867
Light GBM	No	-	-	0,9637	0,0355	0,9620
	-	✓	X	0,9689	0,0349	0,9726
	Yes	X	✓	0,9860	0,0133	0,9853
	-	✓	✓	0,9888	0,0116	0,9893
RUS Boost	No	-	-	0,9579	0,0390	0,9549
	-	✓	X	0,9621	0,0348	0,9590
	Yes	X	✓	0,9786	0,0173	0,9745
	-	✓	✓	0,9795	0,0167	0,9757

Alarm Storm

Modificación del punto de trabajo en MLP Bayes

- Se modifica la relación entre $\bar{c}_0$ y $\bar{c}_1$ en $L^{Bayes}(\mathbf{w}) = \bar{c}_0 \hat{p}_{FA} + \bar{c}_1 \hat{p}_M$

► Ejemplo: $\bar{c}_0 = 1, \bar{c}_1 = 2^k$



Conclusiones

- Complejidad de los datos
 - ▶ Las alarmas de un NOC presentan importantes desafíos: desequilibrio, dimensionalidad, heterogeneidad
 - ★ Técnicas de modelado especializadas

Conclusiones

- Complejidad de los datos

- ▶ Las alarmas de un NOC presentan importantes desafíos: desequilibrio, dimensionalidad, heterogeneidad
 - ★ Técnicas de modelado especializadas

- Información contextual

- ▶ La incorporación de información contextual (estadísticas temporales o patrones de secuencias) mejora las prestaciones de los predictores
 - ★ Ayudan a reflejar la situación actual de la estación y del sistema

Conclusiones

- Complejidad de los datos

- ▶ Las alarmas de un NOC presentan importantes desafíos: desequilibrio, dimensionalidad, heterogeneidad
 - ★ Técnicas de modelado especializadas

- Información contextual

- ▶ La incorporación de información contextual (estadísticas temporales o patrones de secuencias) mejora las prestaciones de los predictores
 - ★ Ayudan a reflejar la situación actual de la estación y del sistema

- Selección del modelo

- ▶ La elección óptima depende del equilibrio operacional entre limitar las falsas alarmas y garantizar una adecuada probabilidad de detección
 - ★ El punto de trabajo “óptimo” depende de los requisitos del centro de operación