

SaludAlpes



# Detección de la Diabetes

FELIPE BEDOYA, CAMILO SALINAS,  
NICOLÁS ORJUELA



# Temas

Problema

Preprocesamiento de los datos

Decision Tree

KNN

Random Forest

Resultados y Recomendaciones

Impacto

Futuras mejoras



AGENDA





# Colombia y la Diabetes

**1.676.885 DE DIABETICOS**

Bajo el supuesto que todos los  
diabeticos reportados no tienen  
complicaciones

Estos le cuestan al sistema salud:

**+524M USD/Año**

# PREPROCESAMIENTO DE LOS DATOS

Técnicas usadas:

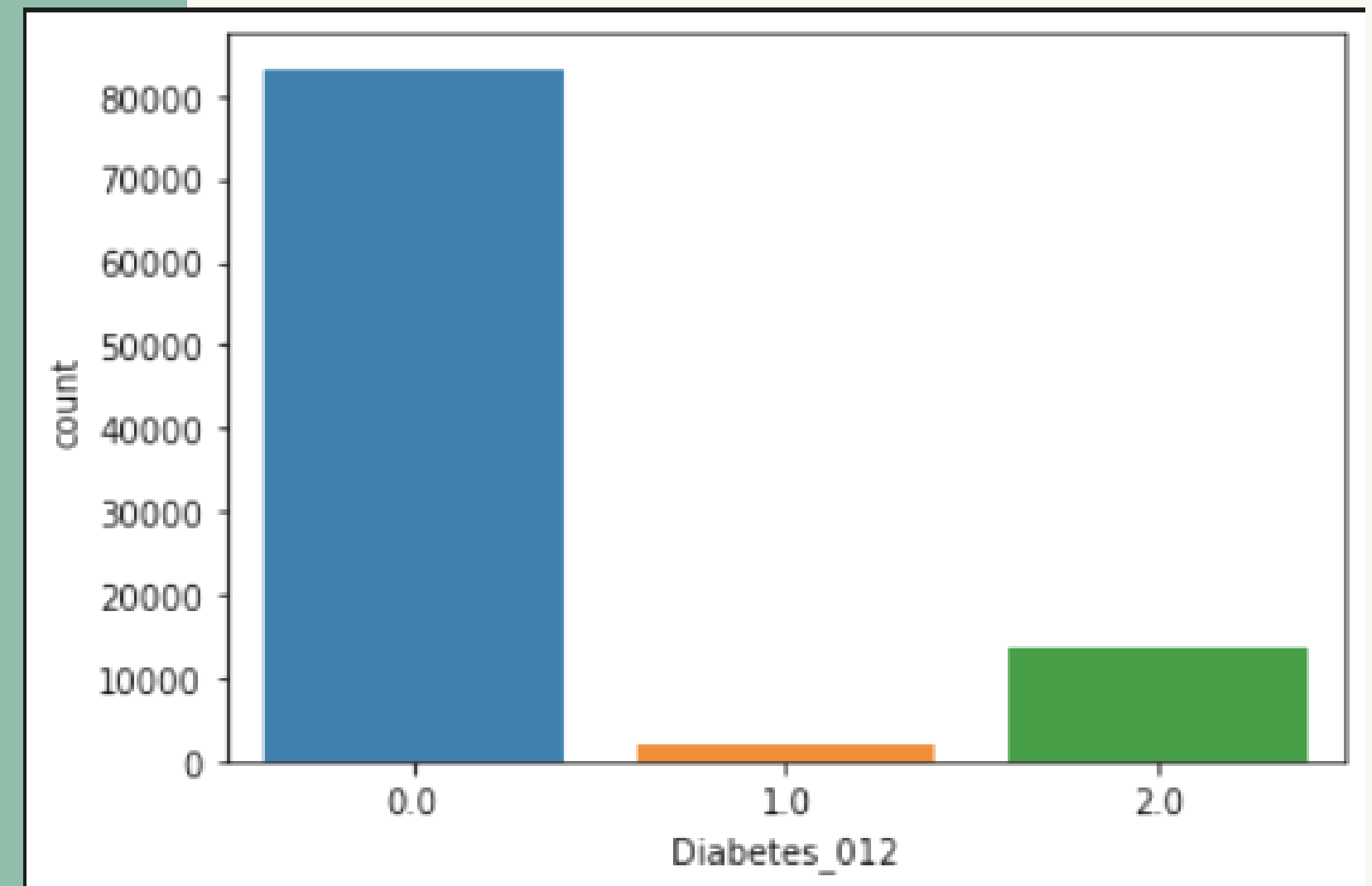
- Limpieza de valores faltantes
- Eliminación de columnas vacías
- Balanceo de los datos
- Método pearson y correlación

KNN

- Normalización de las columnas

Árboles (Decision y Random)

- Estandarización de los datos



# Decision Tree

## Gini

	precision	recall	f1-score	support
0.0	0.85	0.87	0.86	16583
1.0	0.85	0.88	0.86	16554
2.0	0.78	0.73	0.76	16804
accuracy			0.83	49941
macro avg	0.83	0.83	0.83	49941
weighted avg	0.83	0.83	0.83	49941

## Entropía

	precision	recall	f1-score	support
0.0	0.85	0.87	0.86	16583
1.0	0.85	0.88	0.87	16554
2.0	0.78	0.74	0.76	16804
accuracy			0.83	49941
macro avg	0.83	0.83	0.83	49941
weighted avg	0.83	0.83	0.83	49941

ELEGIMOS ENTROPÍA  
Dió mejores resultados que por  
coeficiente de Gini

# KNN

## A PARTIR DE MÉTRICAS DE DISTANCIA

Elegimos los pacientes que más se parecían a diabéticos anteriormente identificados

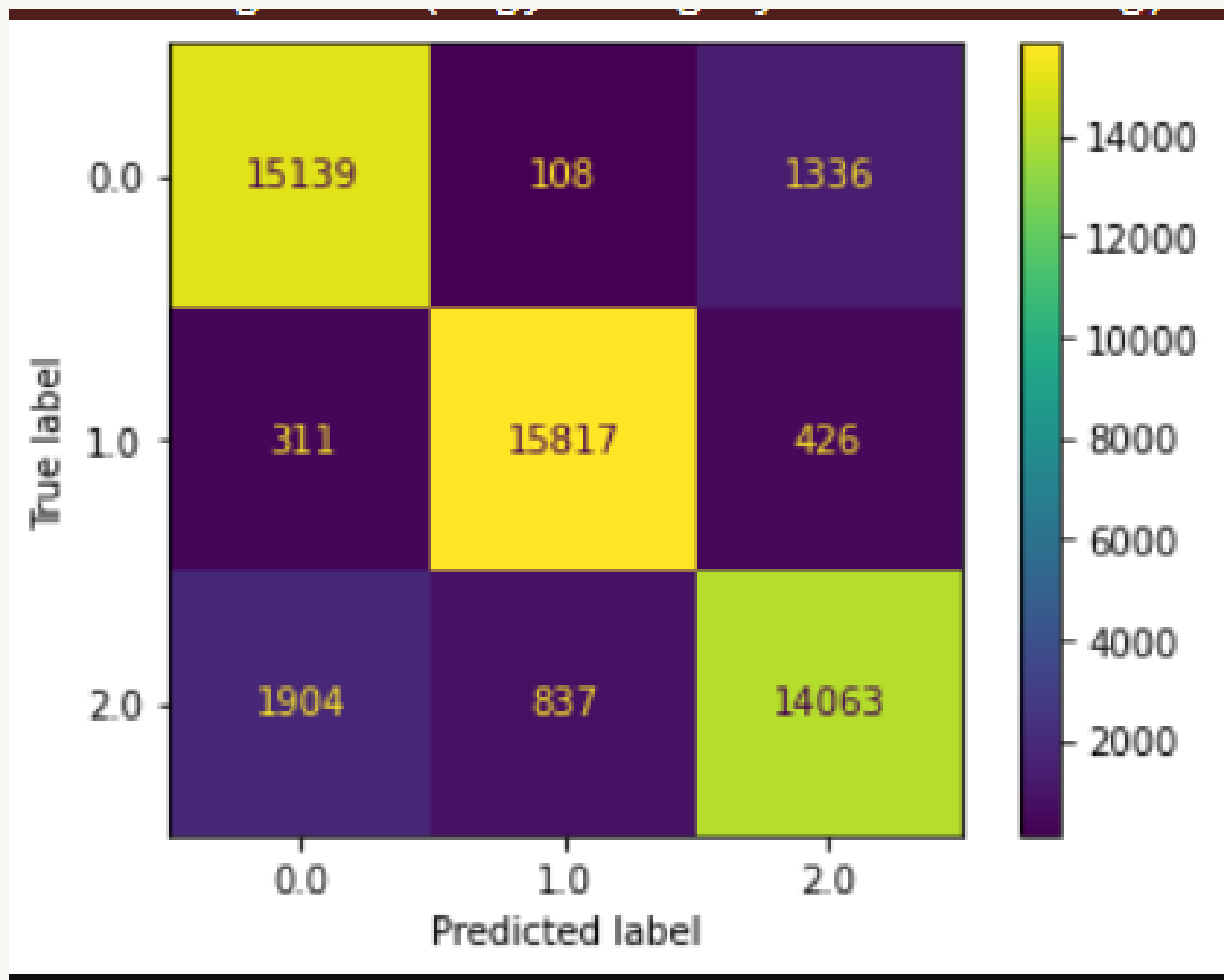
```
Exactitud sobre entrenamiento: 0.81
```

```
Exactitud sobre test: 0.78
```

```
print(classification_report(Y_test, y_pred_test))
```

	precision	recall	f1-score	support
0.0	0.75	0.79	0.77	12575
1.0	0.82	0.87	0.84	12497
2.0	0.76	0.68	0.72	12384
accuracy			0.78	37456
macro avg	0.78	0.78	0.78	37456
weighted avg	0.78	0.78	0.78	37456

# Random Forest



LAS MEJORES PREDICCIONES  
Seccionamos la población a partir de  
comparaciones sobre valores de los  
pacientes. Por ejemplo:

Presión sanguínea < 40

Precisión en todos los casos superior al  
90%



INSIGHTS DE LA INVESTIGACIÓN

# Factores clave para el diagnóstico

se evidencia que hay ciertos factores que tienen alta correlación con el diagnóstico de diabetes, entre los cuales se resalta la salud general del paciente (GenHealth) y la alta presión sanguínea.



# Recomendaciones



## SE LE RECOMIENDA A SALUD ALPES

Utilizar el modelo de Random forest para los diagnósticos

	precision	recall	f1-score	support
0.0	0.87	0.91	0.89	16583
1.0	0.94	0.96	0.95	16554
2.0	0.89	0.84	0.86	16804
accuracy			0.90	49941
macro avg	0.90	0.90	0.90	49941
weighted avg	0.90	0.90	0.90	49941

# Impacto

AL PREDECIR  
TEMPRANAMENTE

## SISTEMA SALUD

Reducción en el costo de los pacientes, para el estado eps's e ips's

## COSTOS OPERACIONALES

SaludAlpes y otras entidades podrían reducir sus costos

## COSTOS PARA LOS PACIENTES

Al tener una detección temprana y acertada se pueden tomar medidas preventivas

# Mejoras



## PROPONEMOS A SALUDALPES LO SIGUIENTE:

Finalmente como recomendación para SaludAlpes, se menciona que la calidad de los datos aunque era buena no era óptima, y esto podría afectar futuros entrenamientos del modelo y diagnósticos.