



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

**VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y TRANSFERENCIA DE
TECNOLÓGICA
CENTRO DE POSGRADOS**

MAESTRÍA EN INGENIERÍA EN SOFTWARE

**TEMA:
“DESARROLLO DE UN MODELO PREDICTIVO PARA LA EVALUACIÓN DEL
RIESGO CREDITICIO EN LA COOPERATIVA DE AHORRO Y CRÉDITO VIRGEN DEL
CISNE.”**

**AUTOR: CHILUIZA MOLINA, OSCAR WLADIMIR
DIRECTORA: MSC. GUEVARA VEGA, CATHY PAMELA**

LATACUNGA, 2021





Planteamiento del Problema

DÉFICIT DE
ANÁLISIS

PROCESO DE
CRÉDITO



DESCONOCIMIENTO

POLÍTICAS

REGLAMENTOS

MANUALES



¿Cómo el modelo de predicción evaluará el riesgo crediticio en la Cooperativa de Ahorro y Crédito Virgen del Cisne?





Objetivos de la Investigación

- **General:**

- Desarrollar un modelo predictivo para la evaluación del riesgo crediticio en la Cooperativa de Ahorro y Crédito Virgen del Cisne.

- **Específicos:**

- Construir Marco Teórico.
- Desarrollar el modelo predictivo.
- Implementar el modelo predictivo.
- Validar el modelo predictivo.



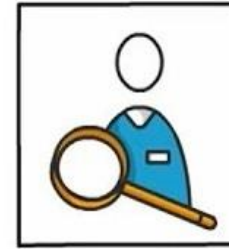
Metodología de Investigación

Investigación aplicada

- Su finalidad es resolver problemas específicos.
- Su objetivo es encontrar una solución.



Niveles de Investigación



Exploratorio



Descriptivo





Marco Teórico

Antecedentes
Históricos

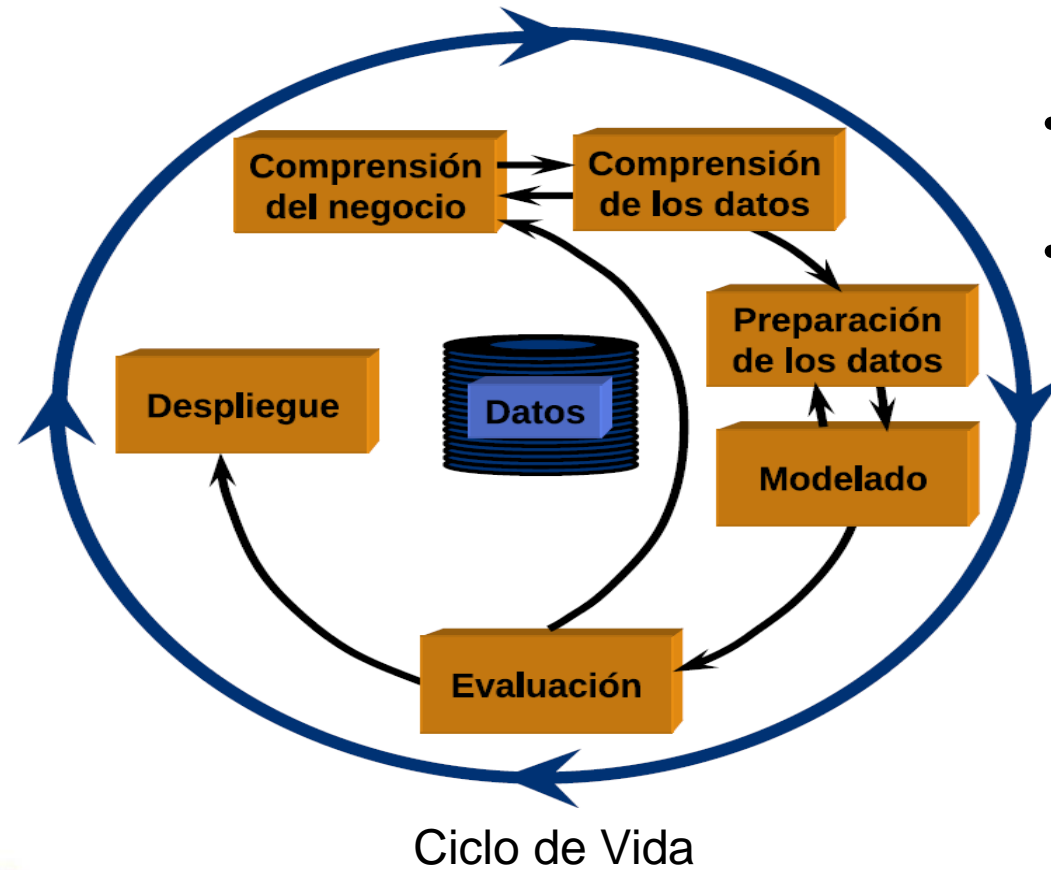
Antecedentes
Referenciales

Antecedentes
Conceptuales

Antecedentes
Contextuales



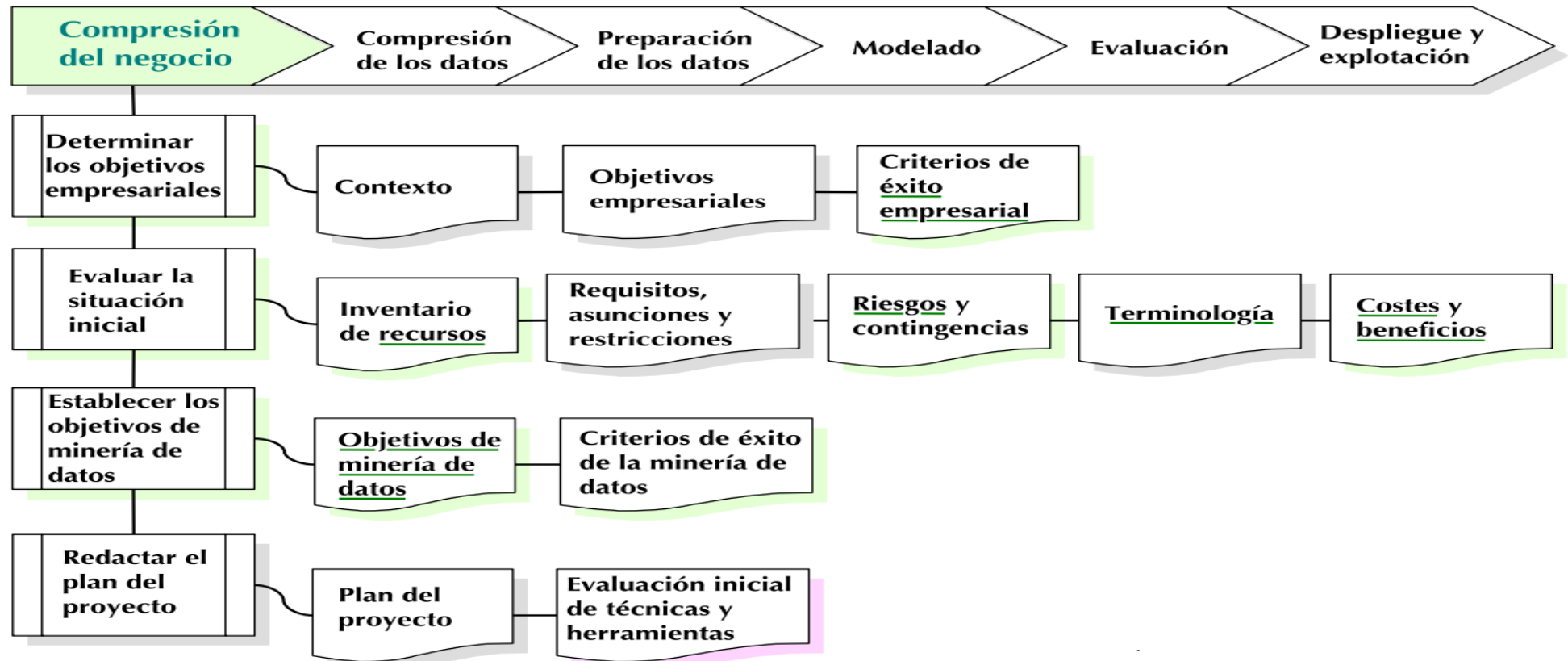
Metodología CRISP-DM



- Metodología para Proyectos de Análisis de Datos.
- Metodología de Minería de Datos basados en fases.



Fase 1: Compresión del negocio





Fase	Tiempo	Equipo
Comprensión del negocio	2 semanas	Jefe Operativo, Investigador
Comprensión de los datos	2 semanas	Departamento de TI, Investigador
Preparación de los datos	3 semanas	Investigador
Modelado	3 semanas	Investigador
Evaluación	2 semanas	Jefe Operativo, Investigador
Implementación o despliegue	1 semana	Departamento de TI, Investigador

Objetivos empresariales

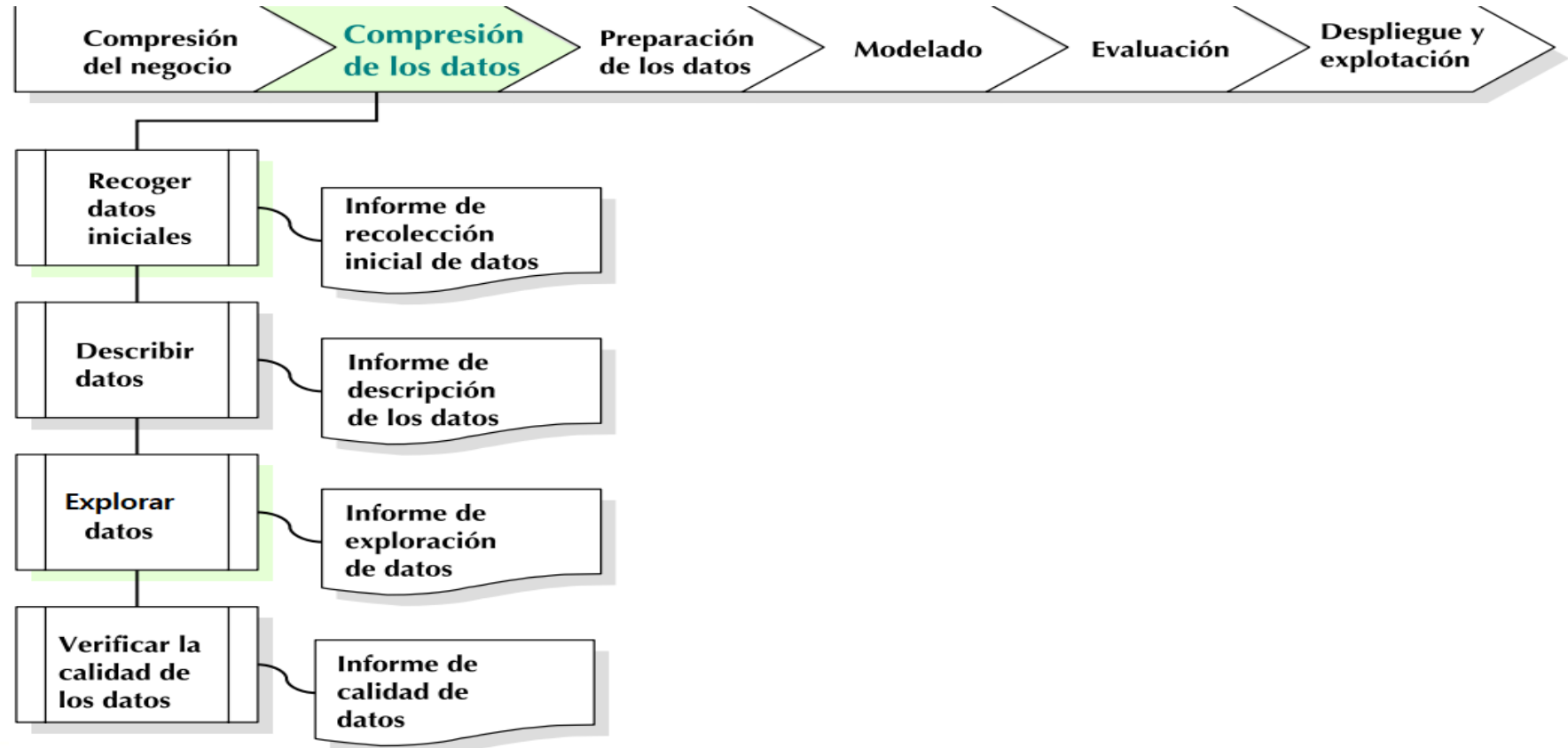
- Mejorar la evaluación de un socio.
- Agilizar el proceso de calificación y entrega de un crédito.
- Minimizar la probabilidad de incumplimiento en los pagos del crédito.

Objetivos de minería de datos

- Generar patrones que ayuden a la evaluación de un cliente.
- Generar patrones que ayuden con la problemática del incremento de mora y baja de rentabilidad.



Fase 2: Compresión de los datos





No	Variable
1	Cargas familiares
2	Destino del Crédito
3	Edad
4	Estado Civil
5	Genero
6	Garantía
7	Monto
8	Nivel Instrucción
9	Préstamos Instituciones
10	Número de créditos
11	Tipo de crédito
12	Tipo de Vivienda
13	Ingresos
14	Egresos

No	Variable
15	Activos
16	Pasivos
17	Tasa de Interés
18	Antigüedad del socio en la institución
19	Antigüedad laboral
20	Zona geográfica
21	Número de Cuotas
22	Frecuencia de Pago
23	Actividad Económica
24	Valor cuota
25	Capacidad de pago
26	Tipo de cliente
27	Calificación del cliente



Fase 3: Preparación de los datos

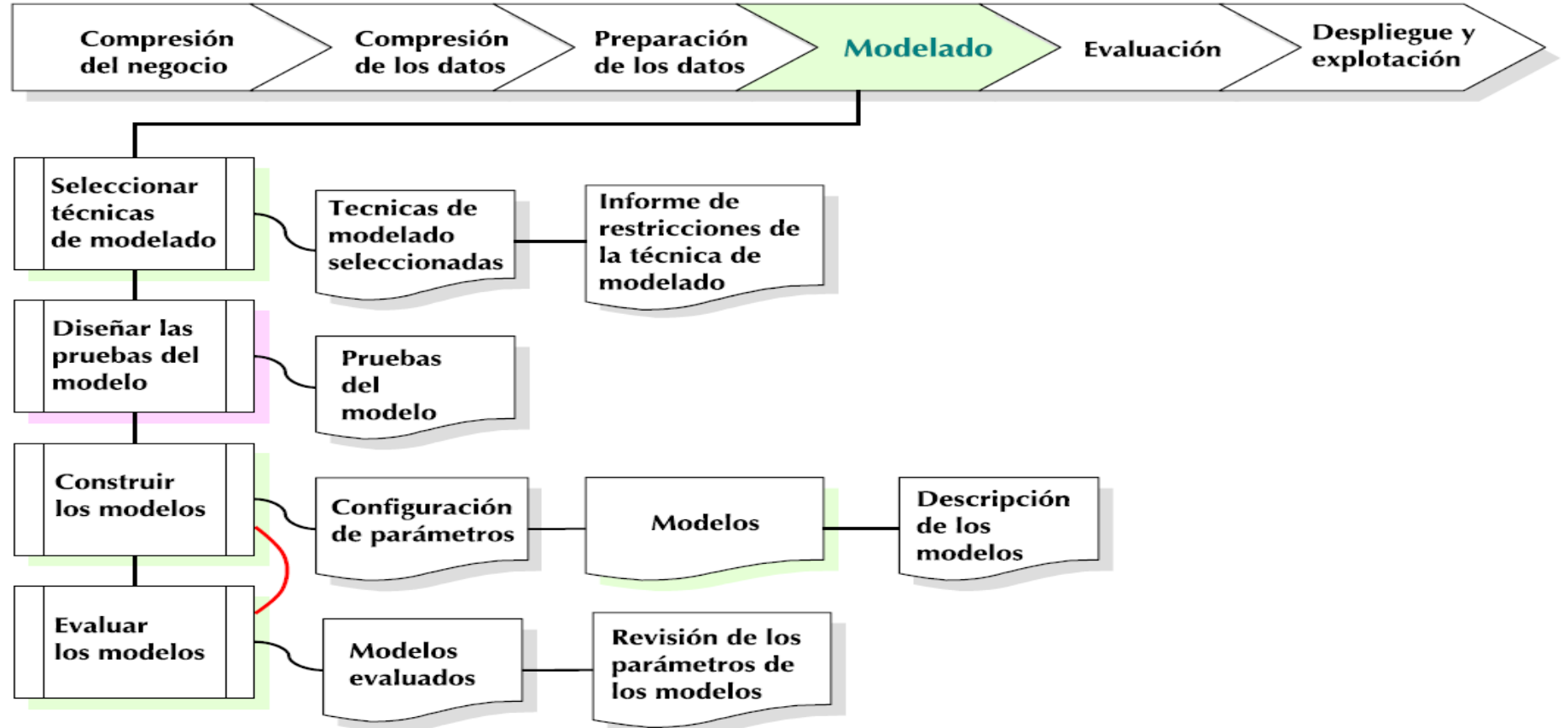




num_cargas	dest_credito	edad	estd_civil	genero	garantia	monto	instruccion	pre_institucionales	num_creditos	tip_credito	tip_vivienda	ingresos	egresos	activos
4	1	35	2	2	0	3550	1	1	2	4	1	840	350	2000
4	1	35	2	2	0	4000	1	1	2	4	1	840	350	2000
1	1	70	2	1	0	1000	1	0	1	4	1	400	130	11600
4	1	31	3	1	0	1500	1	0	1	4	1	450	250	22500
1	1	68	2	1	0	1000	1	0	1	4	1	250	66	24800
3	1	38	3	1	0	2000	1	1	2	4	1	700	291	40500
3	1	38	3	1	0	800	1	1	2	4	1	700	291	40500
0	1	24	1	1	1	5000	2	0	2	4	3	500	140	15000
0	1	24	1	1	1	1500	2	0	2	4	3	500	140	15000
5	9	45	1	2	0	5000	1	1	1	4	1	1300	415	53900
1	1	68	2	1	0	3000	3	0	1	4	2	700	330	76000
4	8	43	2	1	1	15000	3	1	1	2	1	1934	810	127800
3	1	60	3	1	0	1200	1	1	1	4	1	800	460	67000
0	1	23	1	1	1	700	2	0	2	4	3	400	110	10000
0	1	23	1	1	1	400	2	0	2	4	3	400	110	10000
0	1	45	1	2	0	2000	1	0	1	4	1	300	120	12000
2	1	46	1	1	0	1600	1	1	1	4	1	1800	830	20500
1	1	27	2	2	0	3000	2	0	1	4	3	800	120	43000
3	1	54	5	2	0	5000	2	1	1	4	1	1050	235	127000



Fase 4: Modelado





```
from sklearn.model_selection import train_test_split
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.30, random_state = 100)
```

```
import xgboost as xgb  
from sklearn.metrics import roc_auc_score  
from sklearn.metrics import auc, accuracy_score, confusion_matrix, mean_squared_error
```

```
xgtrain = xgb.DMatrix(X_train, label=y_train)
```

```
clf = xgb.XGBClassifier(objective = 'binary:logistic',  
                        max_depth = 150,  
                        min_child_weight = 25,  
                        subsample = 0.85,  
                        colsample_bytree = 0.8,  
                        gamma = 5,  
                        n_jobs = 16,  
                        learning_rate = 0.025,  
                        seed = 1305)
```

```
xgb_param = clf.get_xgb_params()
```

```
#Realizo una predicción  
y_pred = clf.predict(X_test)
```

```
from sklearn.metrics import classification_report  
print(classification_report(y_test, y_pred ))
```

```
#Verifico La matriz de Confusión  
from sklearn.metrics import confusion_matrix
```

```
matriz = confusion_matrix(y_test, y_pred)  
print('Matriz de Confusión:')  
print(matriz)
```

```
Matriz de Confusión:  
[[17226  80]  
 [ 1785 110]]
```

```
from sklearn.model_selection import train_test_split
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.30, random_state = 1)
```

```
#Defino el algoritmo a utilizar
```

```
from sklearn.linear_model import LogisticRegression  
algoritmo = LogisticRegression()
```

```
#Entreno el modelo
```

```
algoritmo.fit(X_train, y_train)
```

```
#Realizo una predicción
```

```
y_pred = algoritmo.predict(X_test)
```

```
from sklearn.metrics import classification_report  
print(classification_report(y_test, y_pred ))
```

```
#Verifico La matriz de Confusión
```

```
from sklearn.metrics import confusion_matrix
```

```
matriz = confusion_matrix(y_test, y_pred)  
print('Matriz de Confusión:')  
print(matriz)
```

```
Matriz de Confusión:  
[[17334  10]  
 [ 1852  5]]
```

```
from sklearn.model_selection import train_test_split
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.30, random_state = 1)
```

```
from sklearn.preprocessing import StandardScaler  
scaler = StandardScaler()  
scaler.fit(X_train)  
X_train = scaler.transform(X_train)  
X_test = scaler.transform(X_test)
```

```
from sklearn.neural_network import MLPClassifier  
mlp=MLPClassifier(hidden_layer_sizes=(10,10,10), max_iter=500, alpha=0.0001,  
                  solver='adam', random_state=21,tol=0.000000001)
```

```
mlp.fit(X_train, y_train)  
y_pred = mlp.predict(X_test)
```

```
from sklearn.metrics import classification_report  
print(classification_report(y_test, y_pred ))
```

```
#Verifico La matriz de Confusión
```

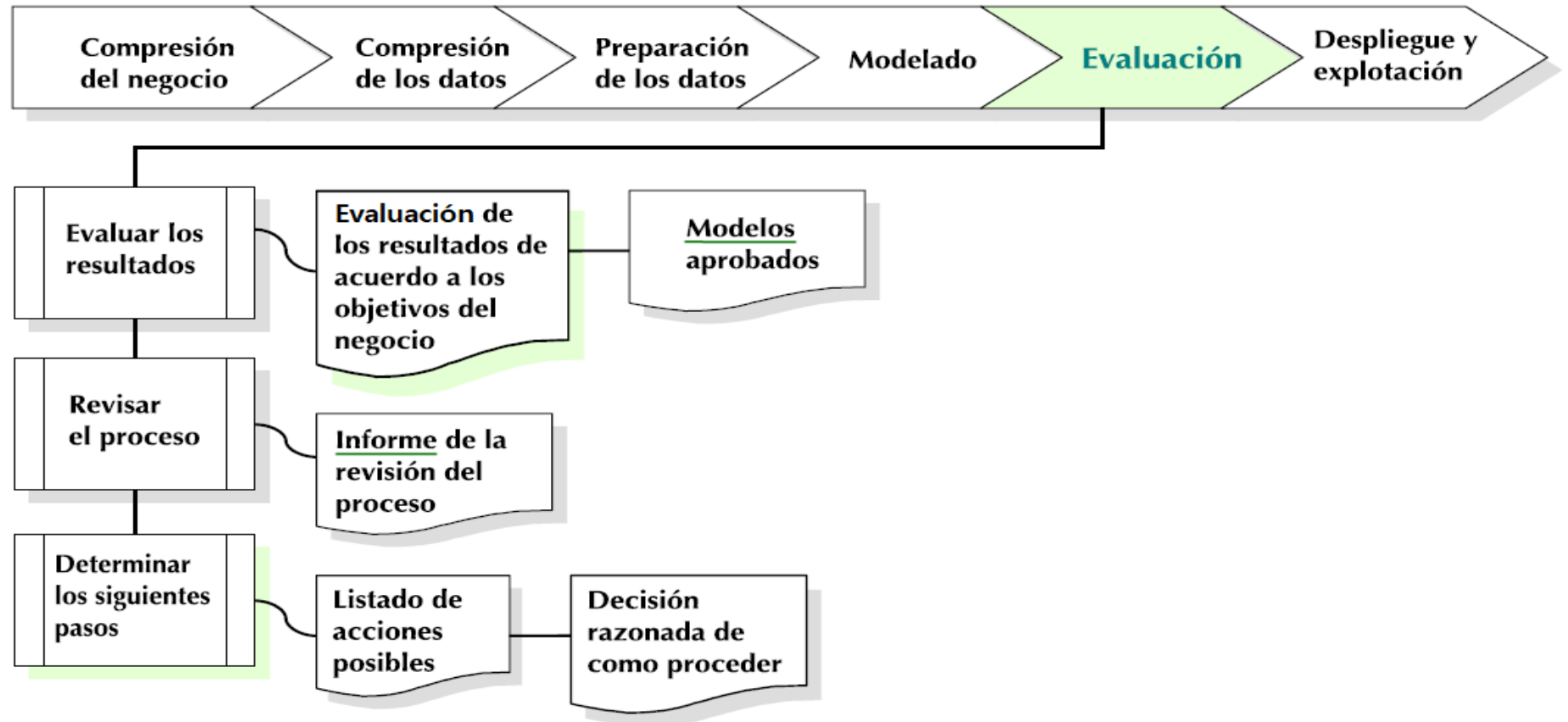
```
from sklearn.metrics import confusion_matrix
```

```
matriz = confusion_matrix(y_test, y_pred)  
print('Matriz de Confusión:')  
print(matriz)
```

```
Matriz de Confusión:  
[[17248  96]  
 [ 1747 110]]
```



Fase 5: Evaluación



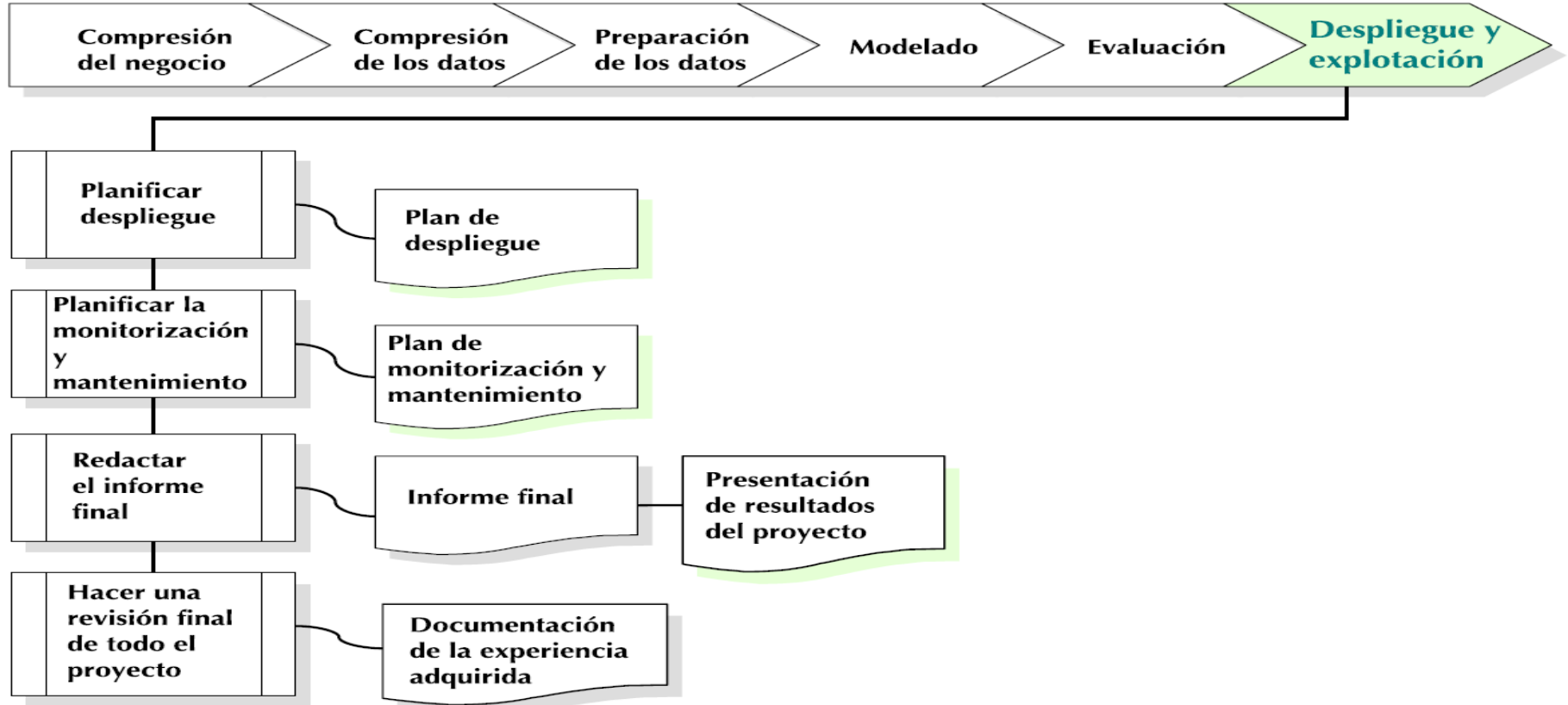


Modelo/Métrica	Exactitud	Tasa de Error	Sensibilidad	Especificidad	Precisión	Predicción Negativa
Random Forest (ML1)	0.90	0.10	0.99	0.06	0.91	0.58
Regresión Logística (ML2)	0.90	0.10	0.99	0.003	0.90	0.33
Red Neuronal (ML3)	0.90	0.10	0.99	0.06	0.91	0.53





Fase 6: Implantación o Despliegue





Virgen del Cisne

COOPERATIVA DE AHORRO Y CRÉDITO

Evaluación del Riesgo de Crédito

Modelo Predictivo - Machine Learning

Carga Familiar	Destino Crédito	Edad	Estado Civil	Género	Garantía	Monto
<input type="text"/>	Selecciona <input type="text"/>	<input type="text"/>	Selecciona <input type="text"/>	Selecciona <input type="text"/>	Selecciona <input type="text"/>	<input type="text"/>
Nivel Instrucción	Préstamos Instituciones	Número créditos	Tipo Crédito	Tipo Vivienda	Ingresos	Egresos
Selecciona <input type="text"/>	Selecciona <input type="text"/>	<input type="text"/>	Selecciona <input type="text"/>	Selecciona <input type="text"/>	<input type="text"/>	<input type="text"/>
Activos	Pasivos	Tasa Interés	Antigüedad Institucional	Antigüedad Laboral	Zona Geografica	Número Cuotas
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	Selecciona <input type="text"/>	<input type="text"/>
Frecuencia Pago	Actividad Económica	Valor cuota	Capacidad Pago	Tipo Cliente		
Selecciona <input type="text"/>	Selecciona <input type="text"/>	<input type="text"/>	Selecciona <input type="text"/>	Selecciona <input type="text"/>		

Predecir Cliente

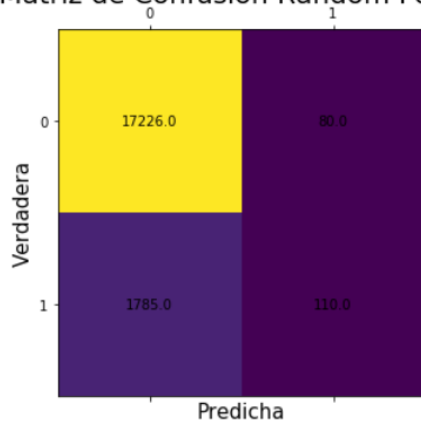
La prediccion del Socio es: --BUENA--





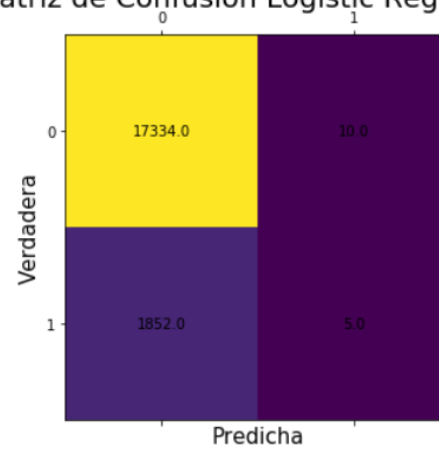
Resultados

Matriz de Confusión Random Forest



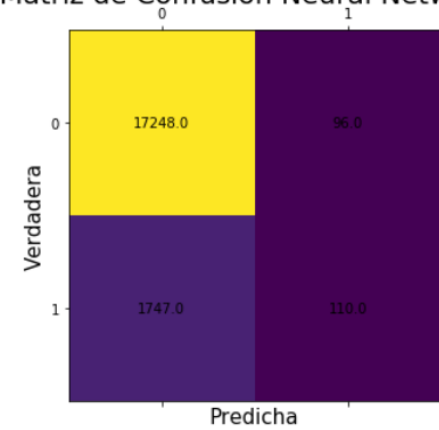
Indica que hay 17336 registros que fueron clasificados correctamente y que 1865 incorrectamente.

Matriz de Confusión Logistic Regression



Indica que hay 17339 registros que fueron clasificados correctamente y que 1862 incorrectamente.

Matriz de Confusión Neural Networks



Indica que hay 17358 registros que fueron clasificados correctamente y que 1843 incorrectamente.





Verificación de Hipótesis

H₀:

Si se desarrolla un modelo predictivo entonces se evaluará el riesgo crediticio optimizando el otorgamiento de créditos en la Cooperativa de Ahorro y Crédito Virgen del Cisne.





Conclusiones

- Se acepta la Hipótesis **H0** planteada con un 95%.
- Técnicas de clasificación binaria, son un método eficaz para la evaluación del riesgo crediticio.
- La metodología CRISP-D genera objetivos que cumplan con las necesidades de la línea del negocio.
- El modelo de Random Forest utilizando el módulo de XGBoots, es el más acertado para predecir si un socio es o no sujeto a crédito.





Recomendaciones

- Analizar los resultados obtenidos en el entrenamiento y test de los modelos en conjunto con Gerencia y Subgerencia.
- Capacitar al personal de TI en temas de Minería de Datos e Inteligencia Artificial.
- Trabajar en conjunto con el área de operaciones, para el análisis de nuevas variables.





Trabajos Futuros

- Predecir el comportamiento del socio cuando decide pre-cancelar una inversión.
- Análisis con Machine Learning en el área de cumplimiento.





ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

“DESARROLLO DE UN MODELO PREDICTIVO PARA LA EVALUACIÓN DEL RIESGO CREDITICIO EN LA COOPERATIVA DE AHORRO Y CRÉDITO VIRGEN DEL CISNE.”

AUTOR: CHILUIZA MOLINA, OSCAR WLADIMIR

LATACUNGA, 2021

