

# 이탈 방지 대작전

3조 부리부리 방범대

김나예 김서진 나성호 서주혁 신동익

# 프로젝트 소개

## 고객 이탈률 예측은 고객 경험을 개선하고 경쟁력을 강화하는 핵심

- 이탈 가능성을 사전 예측하여 고객 불만 해결 및 만족도 향상 가능
- 경쟁사보다 우수한 서비스를 제공, 고객 충성도를 유지하여 장기적인 성장의 발판이 됨

다양한 서비스 제공 기업의 이탈 대응 전략, 서비스 개선안 도출과 효율적인 고객 관리 전략 수립 지원을 위해 **고객 이탈률 예측 모델을 구축**

# 이탈 방지大作战

## 목표

여러 ML 모델 비교를 통해 최적의 고성능 고객 이탈 예측 모델 구축

## 개요

1. 훈련·테스트 데이터 선정
2. ML 성능 비교
3. 예측 모델 설계·구축

# 데이터 전처리

## Kaggle 머신러닝 웹 커뮤니티의 **Telecom Churn Dataset** 채택

State	# Account length	# Area code	International plan	Voice mail plan	# Number vmail me...	# Total day minutes
State	Account length	Area code	International plan	Voice mail plan	Number vmail messages	Total day minutes
AZ	3%		true	true		
WA	3%		0 0%	0 0%		
Other (630)	94%		false	false		
	0 232	408 510	0 0%	0 0%	0 51	25.9 3
LA	117	408	No	No	0	184.5
IN	65	415	No	No	0	129.1

1. 불필요한 feature 제외 (['State', 'Area code' ])
2. 범주형 데이터 Label Encoding
3. 정규화

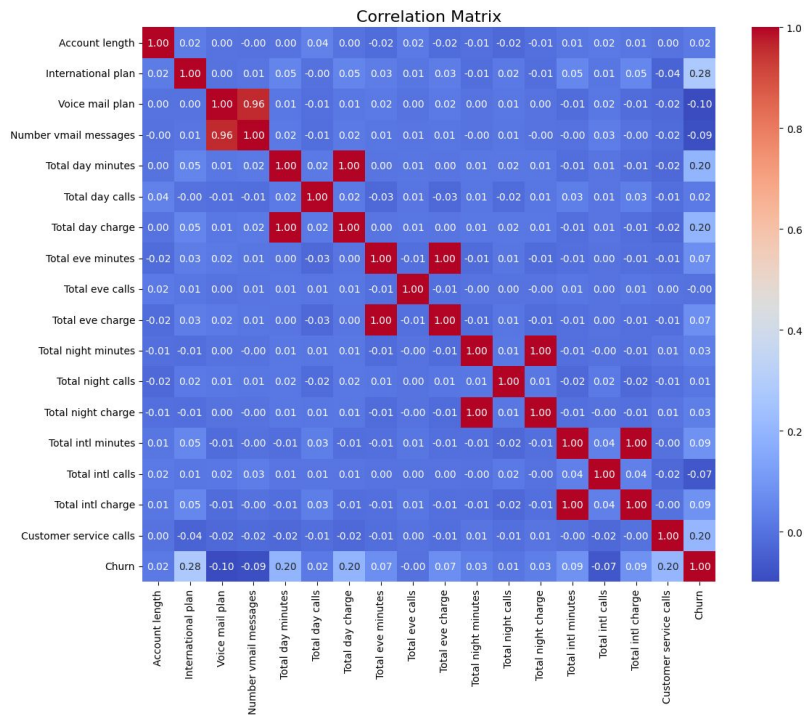
# 데이터 전처리

## Column 설명

Number	Column	Meaning	Example
1	International plan	국제 전화 요금제 가입 여부	"Yes" (가입), "No" (미가입)
2	Voice mail plan	음성사서함 요금제 가입 여부	"Yes" (가입), "No" (미가입)
3	Total day minutes	주간(낮 시간) 동안 사용한 총 통화 시간(분)	265.1 → 265.1분 통화
4	Total eve minutes	저녁 시간 동안 사용한 총 통화 시간(분)	197.4 → 197.4분 통화
5	Total intl minutes	국제 통화에 사용된 총 시간(분)	10.0 → 10분 사용
6	Customer service calls	고객 센터에 전화한 횟수	1 → 1회 전화
7	Churn	고객 이탈 여부	True (이탈), False (유지)

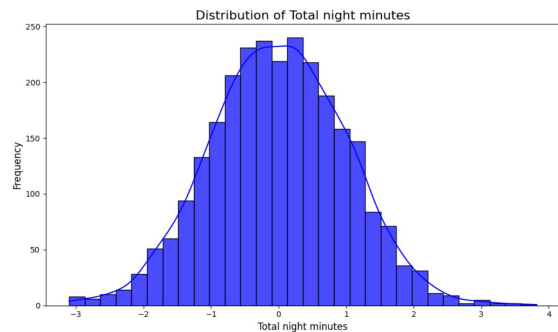
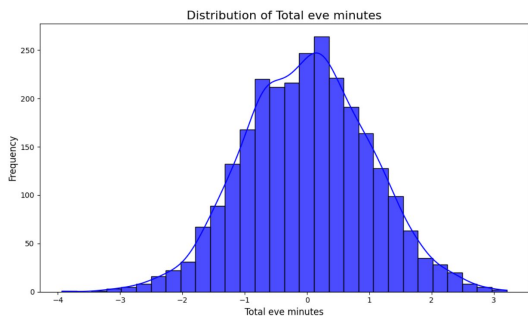
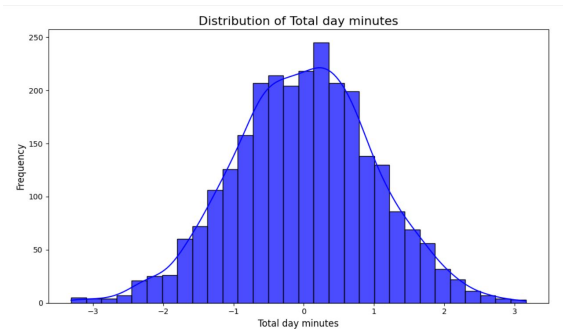
# EDA

## Correlation Matrix



# EDA

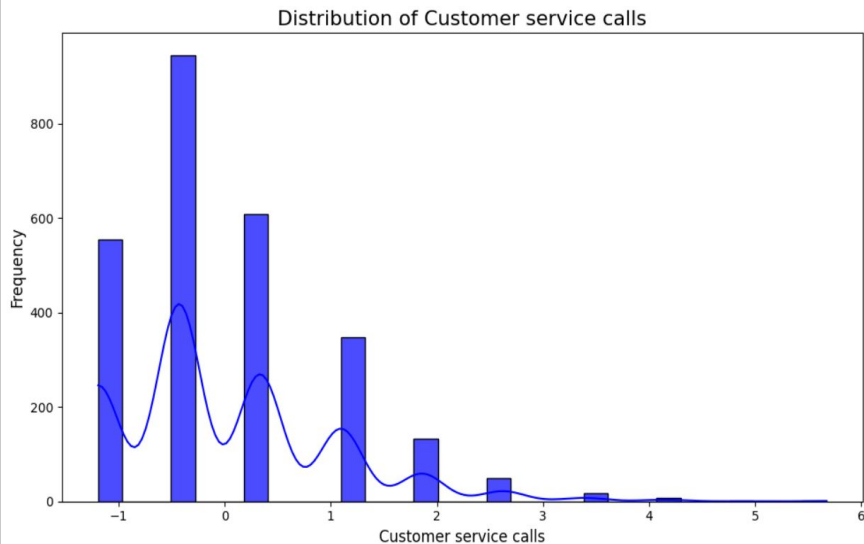
## 주간/야간/심야 통화 이용률



# EDA

## 고객센터 상담 비율

고객센터 통화량



고객센터 통화량 대비 이탈률

Customer service calls

-1.191955 14.234234

-0.429172 10.476190

0.333610 10.197368

1.096392 10.632184

1.859175 48.120301

2.621957 59.183673

3.384740 58.823529

4.147522 62.500000

4.910305 100.000000

5.673087 100.000000

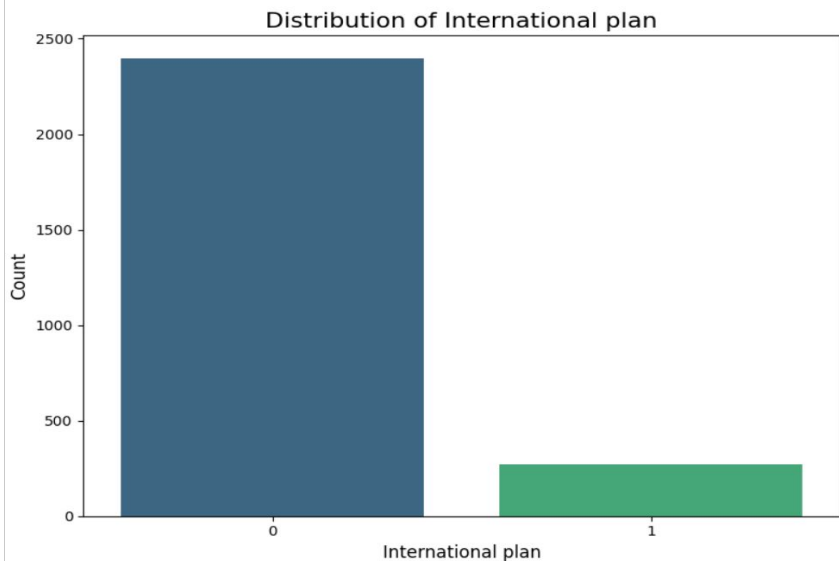
Name: Churn, dtype: float64



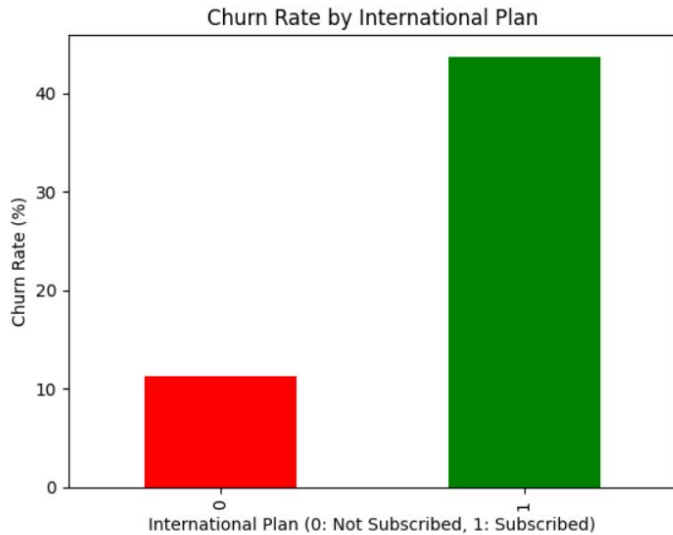
# EDA

## 국제전화/음성사서함 가입 비율

국제전화 가입 비율



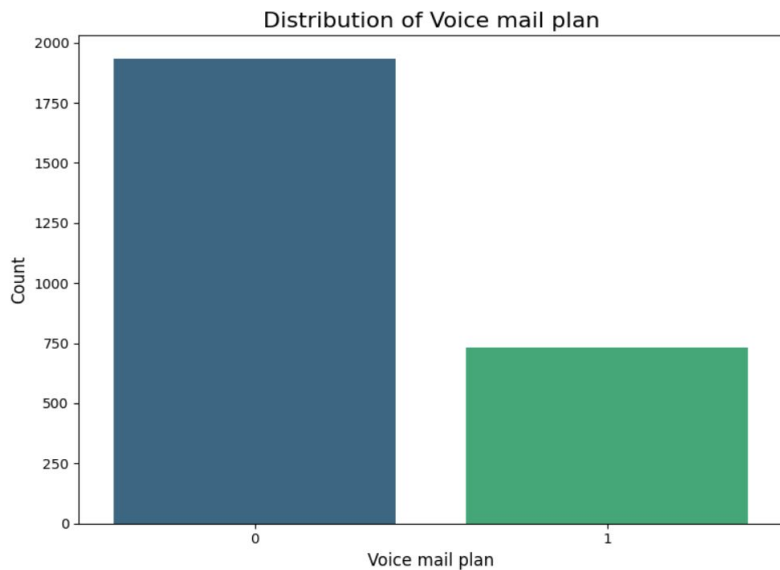
국제전화 가입 비율 대비 이탈률



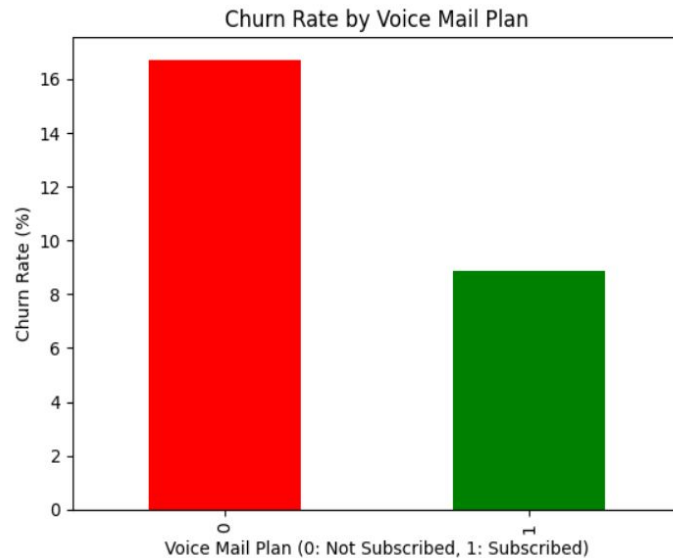
# EDA

## 국제전화/음성사서함 가입 비율

음성사서함 가입 비율

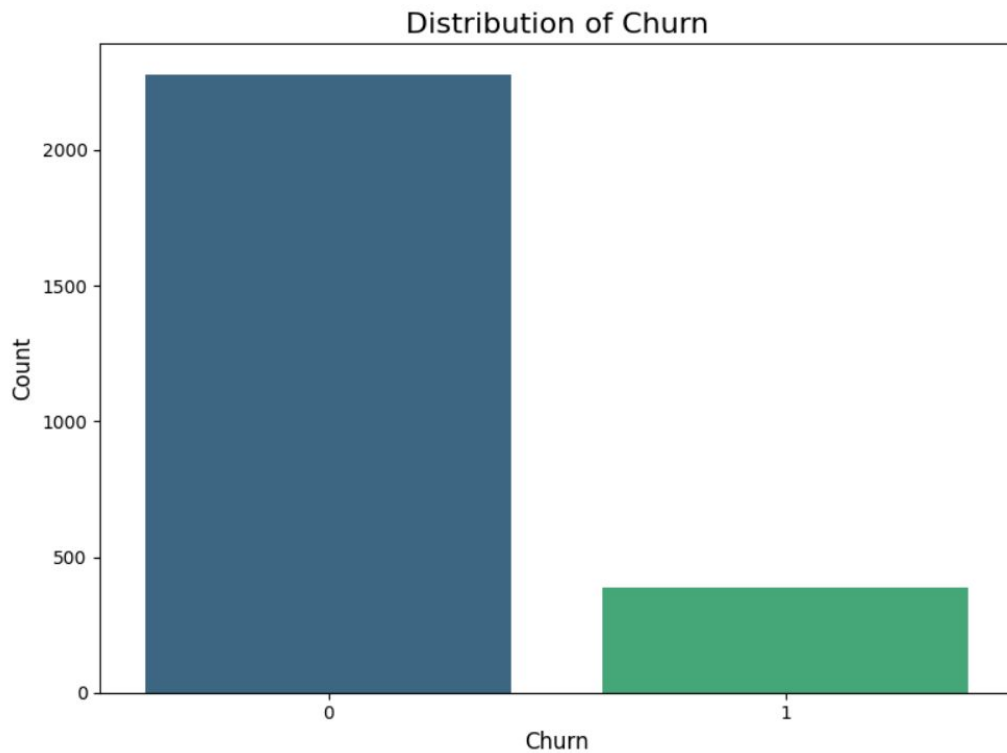


음성사서함 가입 비율 대비 이탈률



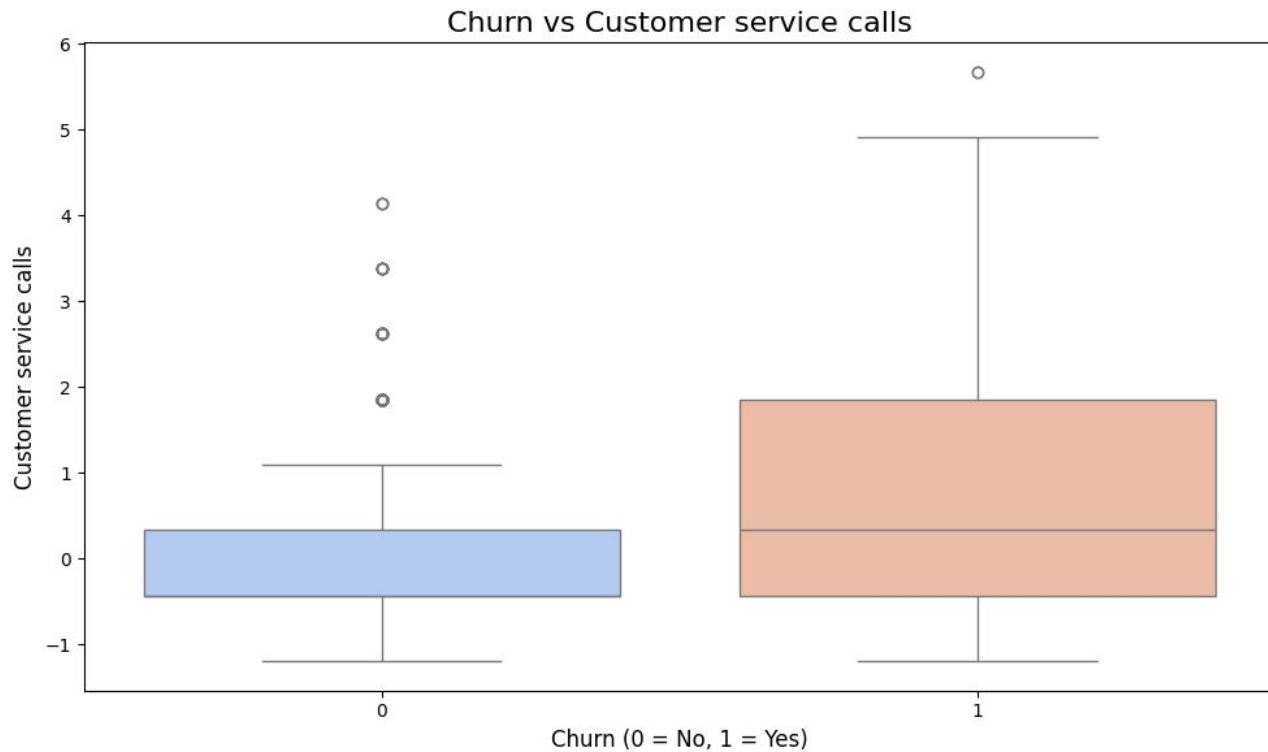
# EDA

## 이탈률



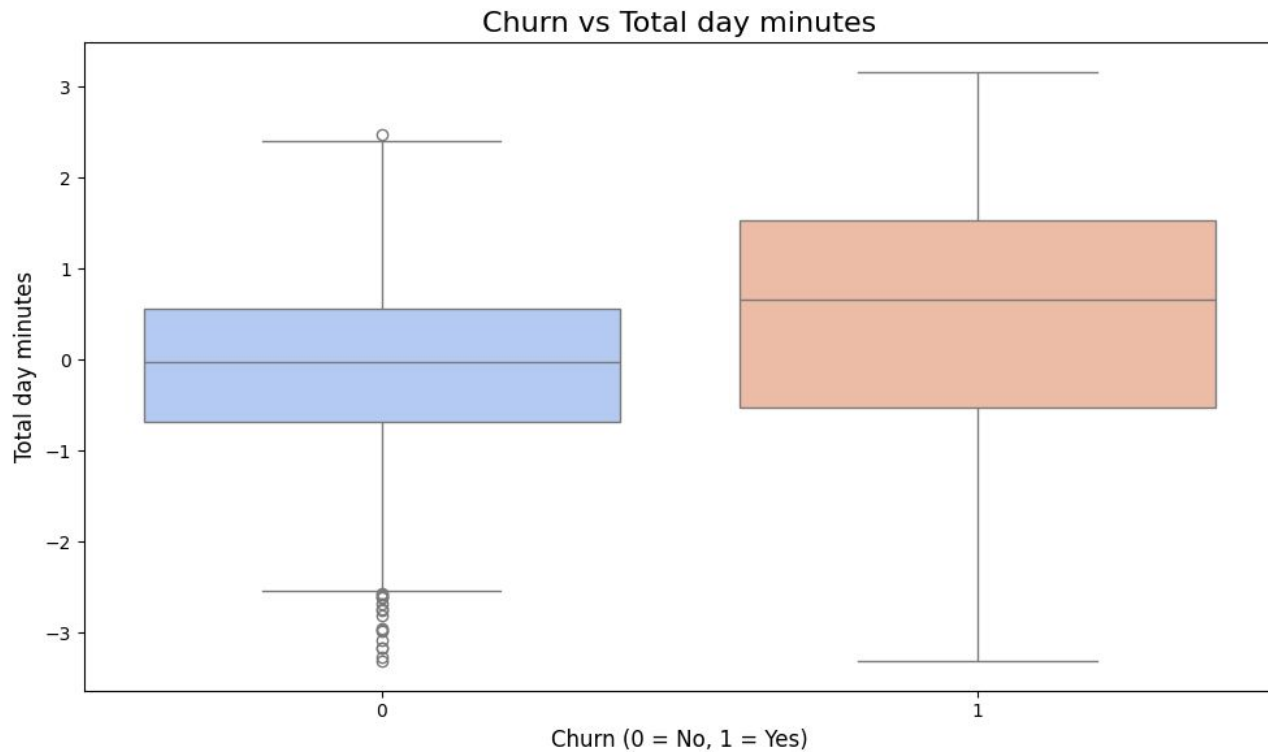
# EDA

## 고객센터 통화량 대비 이탈률 & 이상치



# EDA

## 통화량 대비 고객 이탈률



# Machine Learning

## 모델별 평가지표

Model	Accuracy	Precision	Recall	F1 Score	ROC AUC
KNN	0.885	0.828	0.304	0.444	0.646
Disition Tree	0.913	0.680	0.737	0.707	0.840
XGBoost	0.957	0.934	0.747	0.830	0.911
Random Forest	0.958	0.972	0.737	0.838	0.924

- XGBoost와 Random Forest 모델이 가장 높은 성능을 보임

# 최종 모델 선정 (XGBoost)

## 1. 균형 잡힌 성능

- Recall이 Random Forest보다 높아 이탈 고객을 더 많이 예측할 수 있습니다.
- F1 Score도 높은 수준을 유지하며 정확도와 재현율 간 균형을 이룹니다.

## 2. 과적합 방지

- Random Forest는 Precision은 높았지만 과적합 경향이 보였습니다.
- XGBoost는 안정적이면서도 일반화 성능이 뛰어났습니다.

## 3. 해석 가능성

- XGBoost를 통해 Feature Importance를 분석하면 이탈 요인을 명확히 파악할 수 있습니다.

# Deep Learning

## RNN Model

### 1. RNN 모델 생성

```
rnn = Sequential([  
    SimpleRNN(32, activation='relu', input_shape=(X_train_scaled.shape[1], 1)),  
    Dropout(0.5),  
    Dense(1, activation='sigmoid')  
])
```

### 2. 과적합 방지

```
early_stopping = EarlyStopping(monitor='val_loss', patience=5, restore_best_weights=True)
```



# Deep Learning

## RNN Model

### 3. 훈련 결과 (Epoch : 34)

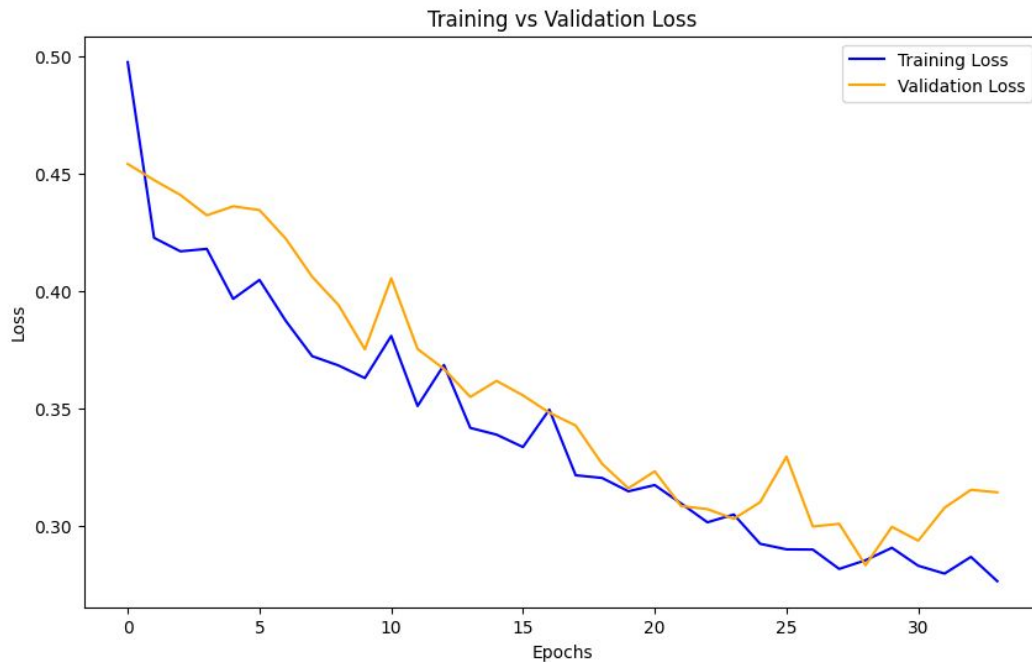
```
Epoch 10/100
67/67 ██████████ 1s 7ms/step - accuracy: 0.8647 - loss: 0.3646 - val_accuracy: 0.8333 - val_loss: 0.3752
Epoch 11/100
67/67 ██████████ 1s 7ms/step - accuracy: 0.8678 - loss: 0.3533 - val_accuracy: 0.8221 - val_loss: 0.4054
Epoch 12/100
67/67 ██████████ 1s 11ms/step - accuracy: 0.8749 - loss: 0.3410 - val_accuracy: 0.8352 - val_loss: 0.3753
Epoch 13/100
...
Epoch 33/100
67/67 ██████████ 1s 7ms/step - accuracy: 0.8823 - loss: 0.2746 - val_accuracy: 0.8689 - val_loss: 0.3154
Epoch 34/100
67/67 ██████████ 1s 7ms/step - accuracy: 0.9063 - loss: 0.2858 - val_accuracy: 0.8708 - val_loss: 0.3143
```

Output is truncated. View as a [scrollable element](#) or open in a [text editor](#). Adjust cell output [settings](#)...

# Deep Learning

## RNN Model

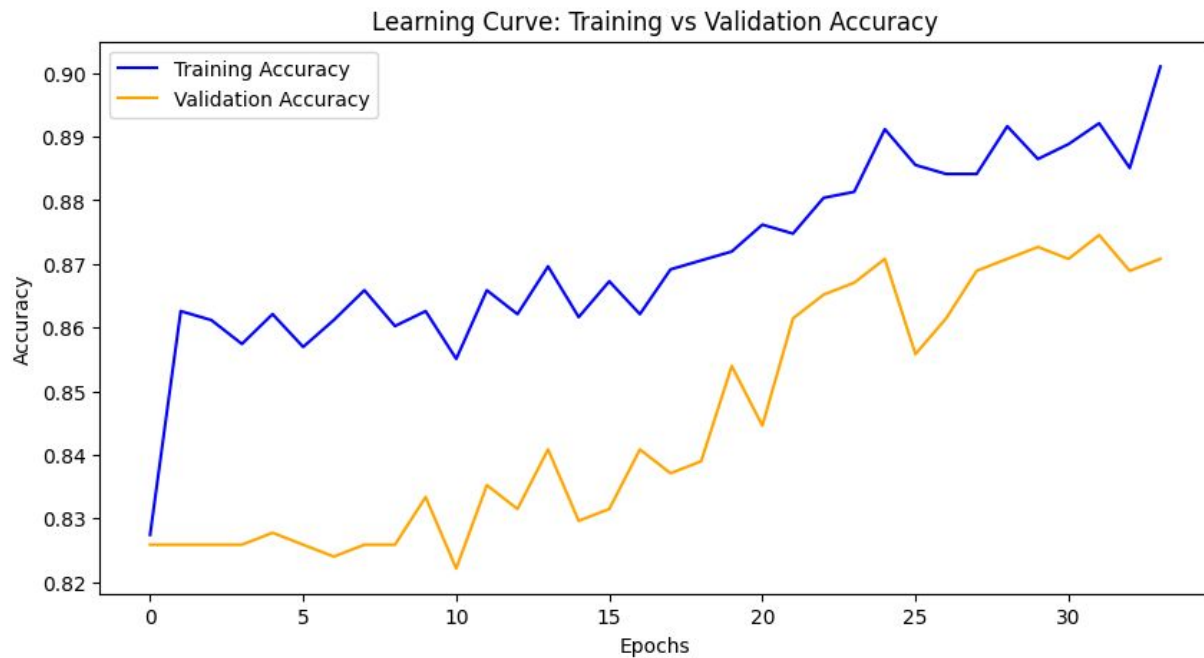
### 4. Loss Curve



# Deep Learning

## RNN Model

### 5. Learning Curve



# Deep Learning

## RNN Model

### 6. 평가

RNN 모델 평가 결과:

	Accuracy	Precision	Recall	F1 Score	ROC AUC
RNN	0.889055	0.672131	0.431579	0.525641	0.698307

# RNN 모델의 한계

## 1. 성능 부족

- a. 정확도(Accuracy)가 88.9%로 나쁘지 않으나, Recall이 43.2%로 매우 낮음
- b. 이는 이탈 고객을 제대로 예측하지 못하는 한계가 있음을 의미

## 2. F1 Score

- a. F1 Score가 0.526으로 낮아 Precision과 Recall의 균형이 부족함

## 3. ROC-AUC

- a. ROC-AUC 점수가 0.698로, 다른 머신러닝 모델(XGBoost, Random Forest)에 비해 식별 능력이 떨어짐

# 결론

## XGBoost와의 비교

1. XGBoost는 모든 성능 지표에서 RNN보다 우수한 결과를 보임
2. 특히 Recall과 ROC-AUC에서 큰 차이를 보이며, 이탈 고객 예측에 더 효과적으로 판단됨

## 결과

RNN은 딥러닝 모델로 적용되었지만, 성능과 효율성 면에서 부족하다고 판단함  
따라서 **XGBoost를 최종 모델로 선정**

# 구축한 모델로 예측 해보기

[고객 이탈률 예측 웹](#)

## Customer Churn Prediction

Enter the customer details to predict churn.

State (e.g., CA, TX, NY)

Account length

 - +

Area code

 - +

International plan

 ▼

Voice mail plan

 ▼

감사합니다