

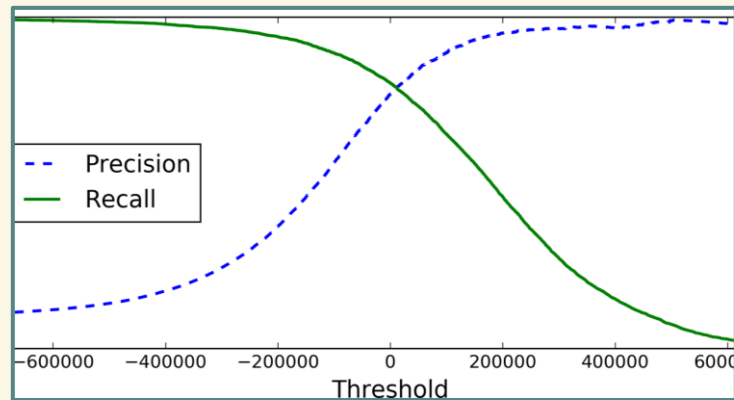


# Метрика **Recall** в задачах машинного обучения

Серкибаев Аким | Э-2209

# Что такое Recall?

**Recall (полнота)** – это доля истинно положительных ответов (True Positives), которые модель верно «нашла» среди всех действительно положительных объектов. Используется, когда важнее всего не упустить положительные примеры (например, при обнаружении заболеваний).



# Формула Recall

$$Recall = \frac{TP}{TP + FN}$$

- **TP (True Positives)** – число истинно положительных ответов.
- **FN (False Negatives)** – число ложноотрицательных ответов.
- Чем выше значение **Recall**, тем меньше пропускается важных (положительных) примеров.

# Матрица ошибок (Confusion Matrix)

		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

- **True Positive (TP):** модель спрогнозировала «положительно», и это верно.
- **False Positive (FP):** модель спрогнозировала «положительно», но на самом деле «отрицательно».
- **False Negative (FN):** модель спрогнозировала «отрицательно», но на самом деле «положительно».
- **True Negative (TN):** модель спрогнозировала «отрицательно», и это верно.
- **Recall** отражает отношение **TP** к сумме **TP+FN**.

# Сравнение Recall с Precision

**Precision** (точность) отвечает за то, сколько истинно положительных ответов среди всех, обозначенных моделью как положительные.

**Recall** (полнота) говорит о том, какой процент положительных ответов из всех положительных был найден.

В реальных задачах баланс между **Precision** и **Recall** подбирается с учётом конкретных требований (**Precision-Recall trade-off**).

$$Precision = \frac{TP}{TP+FP}$$

# Связь Recall с другими метриками

F1-мера объединяет Precision и Recall в единое число:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

- Гармоническое среднее сильно штрафует дисбаланс между **Precision** и **Recall**.
- Высокое значение **Recall** не всегда означает высокое качество модели в целом. Важно анализировать **Recall** совместно с **Precision**, **Accuracy**, **ROC AUC** и др.

# Пример применения Recall



## Медицина

Предположим, что необходимо выявлять у пациентов наличие серьёзного заболевания, и крайне важно не упустить ни одного случая. Высокий **Recall** позволит свести к минимуму риск нераспознанных больных, однако при этом может увеличиться число ложных положительных результатов, что приведёт к лишним анализам и стрессу для здоровых людей.



## Выявление спам-сообщений

В системе фильтрации электронной почты крайне нежелательно, чтобы спам просачивался во «Входящие», поэтому высокий **Recall** способствует задержанию подавляющего большинства нежелательных писем. Однако чрезмерный акцент на **Recall** потенциально увеличит количество полезных писем, ошибочно попавших в «Спам», что говорит о необходимости баланса с **Precision**.



## Детекция финансовых махинаций

Банки и платёжные системы стремятся выявлять подозрительные транзакции, чтобы предотвратить мошенничество, и у них крайне критичен каждый пропущенный случай. При этом слишком высокий **Recall** может вызвать рост ложных блокировок для добросовестных клиентов, требуя тонкого компромисса между безопасностью и удобством.

# Ключевые выводы



**Recall** ориентирован на выявление максимального числа истинно положительных случаев



Полезен при высоких требованиях к обнаружению «позитивных» объектов (например, угроз кибербезопасности)

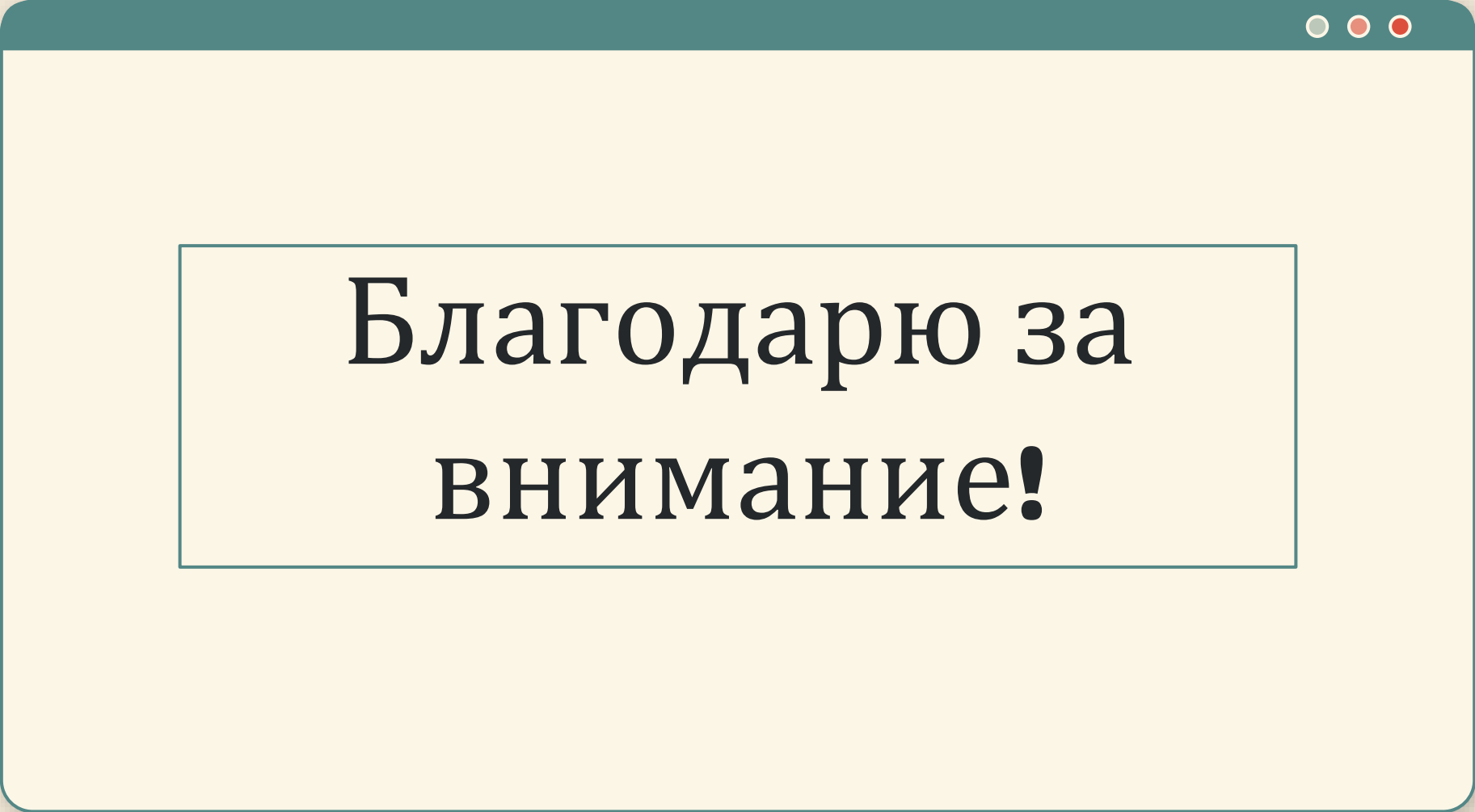


Слишком высокий **Recall** может приводить к росту ложноположительных срабатываний, что не всегда допустимо



Оптимальное значение выбирается с учётом специфики задачи и возможных рисков





Благодарю за  
внимание!



# Метрика **Recall** в задачах машинного обучения

Серкибаев Аким | Э-2209