

# Применение графов для классификации финансовых транзакций

Выполнил:

студент группы М80-404 Сорокин Д.М.

Руководитель:

доцент каф. 804 Соболев В.Р.

# Постановка задачи

Проанализировать финансовые транзакции с целью предотвращения мошеннических операций. Разбивается на две подзадачи:

- Реализовать классификатор способный отличить мошеннические транзакции от не мошеннических
- Построить графы на основе финансовых переводов. На графах посчитать новые признаки (расстояние до ближайшего мошенника, количество мошенников в круге радиуса  $n$ ), добавить их в модель и проверить значимость

# Данные

	step	type	amount	nameOrig	oldbalanceOrig	newbalanceOrig	nameDest	oldbalanceDest	newbalanceDest	isFraud
0	1	PAYMENT	9839.64	C1231006815	170136.00	160296.36	M1979787155	0.0	0.0	0
1	1	TRANSFER	181.00	C1305486145	181.00	0.00	C553264065	0.0	0.0	1
2	1	CASH_OUT	181.00	C840083671	181.00	0.00	C38997010	21182.0	0.0	1
3	1	PAYMENT	9478.39	C1671590089	116494.00	107015.61	M58488213	0.0	0.0	0
4	1	PAYMENT	3454.08	C686349795	9031.96	5577.88	M1831010686	0.0	0.0	0

- **step**                      аналог времени. 1 step = 1 час (всего 744 = 30 дней)
- **type**                     тип транзакции (CASH-IN, CASH-OUT, DEBIT, PAYMENT, TRANSFER)
- **amount**                 сумма перевода
- **nameOrig**                ID пользователя-отправителя
- **oldbalanceOrig**        баланс отправителя до транзакции
- **newbalanceOrig**        баланс отправителя после транзакции
- **nameDest**                ID пользователя-получателя
- **oldbalanceDest**        баланс получателя до транзакции. ID, начинающийся с буквы **M** - Merchant (магазин). В этом случае информация отсутствует
- **newbalanceDest**        баланс получателя после транзакции
- **isFraud**                 пометка о мошеннической транзакции

# Визуализация с помощью графов

4

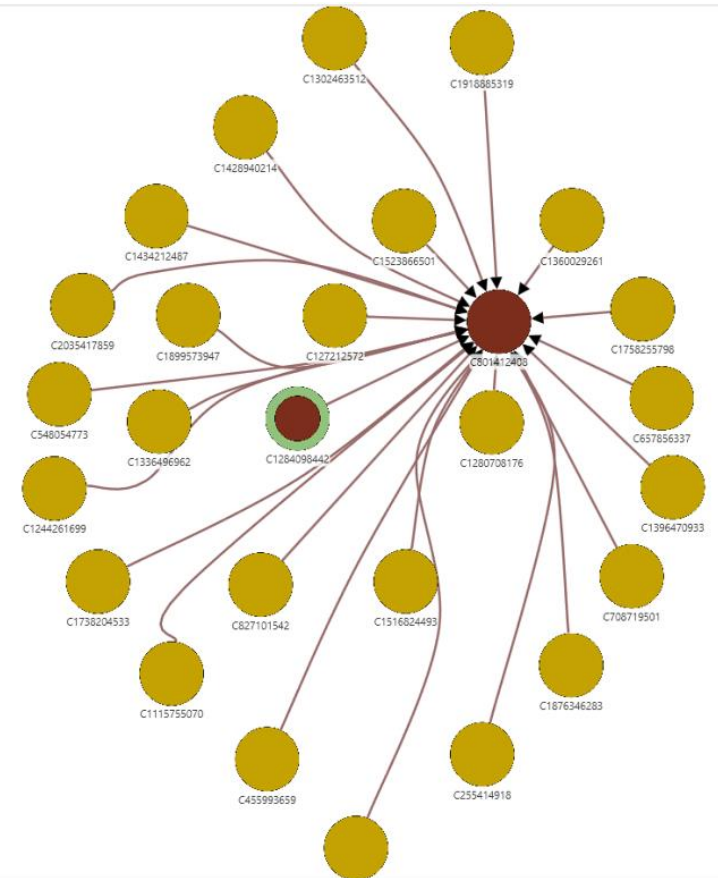


# Предварительный анализ

Название	Значение
Всего транзакций	6362620
Мошеннических	8213
Ср. сумма перевода мошеннической транзакции	1467967.29
Ср. сумма перевода немошеннической транзакции	178197.04
Уникальных клиентов	4777844
Количество магазинов	2151495
Уникальных магазинов	2150401
Мошеннических транзакций с магазинами	0
Клиенты, взаимодействующие друг с другом более 1 раза	0

Название	Значение
CASH-IN (frauds)	1399284 (0)
CASH-OUT	2237500 (4116)
DEBIT	41432 (0)
PAYMENT (=кол-во магазинов)	2151495 (0)
TRANSFER	532909 (4097)
Максимальная цепочка транзакций	11
Максимальное количество приема одним клиентов	118
Максимальное количество отправлений одним клиентов	31

## Типичная ситуация для мошеннических транзакций



# Проектирование признаков

	step	type	amount	oldbalanceOrg	isFraud	hour	newSender	newReceiver	merchant	fraudsEarly	LTS	LTR	IZoB
	0	1	4	181.00	181.0	1	1	1	1	0	0	0	1
	1	1	1	181.00	181.0	1	1	1	1	0	0	0	1
	2	1	3	7107.77	183195.0	0	1	1	1	1	0	0	0
	3	1	3	671.64	15123.0	0	1	1	1	1	0	0	0
	4	1	3	1373.43	13854.0	0	1	1	1	1	0	0	0

Выборка сокращена до **461382** транзакций. Мошеннические сохранены в исходном объеме.

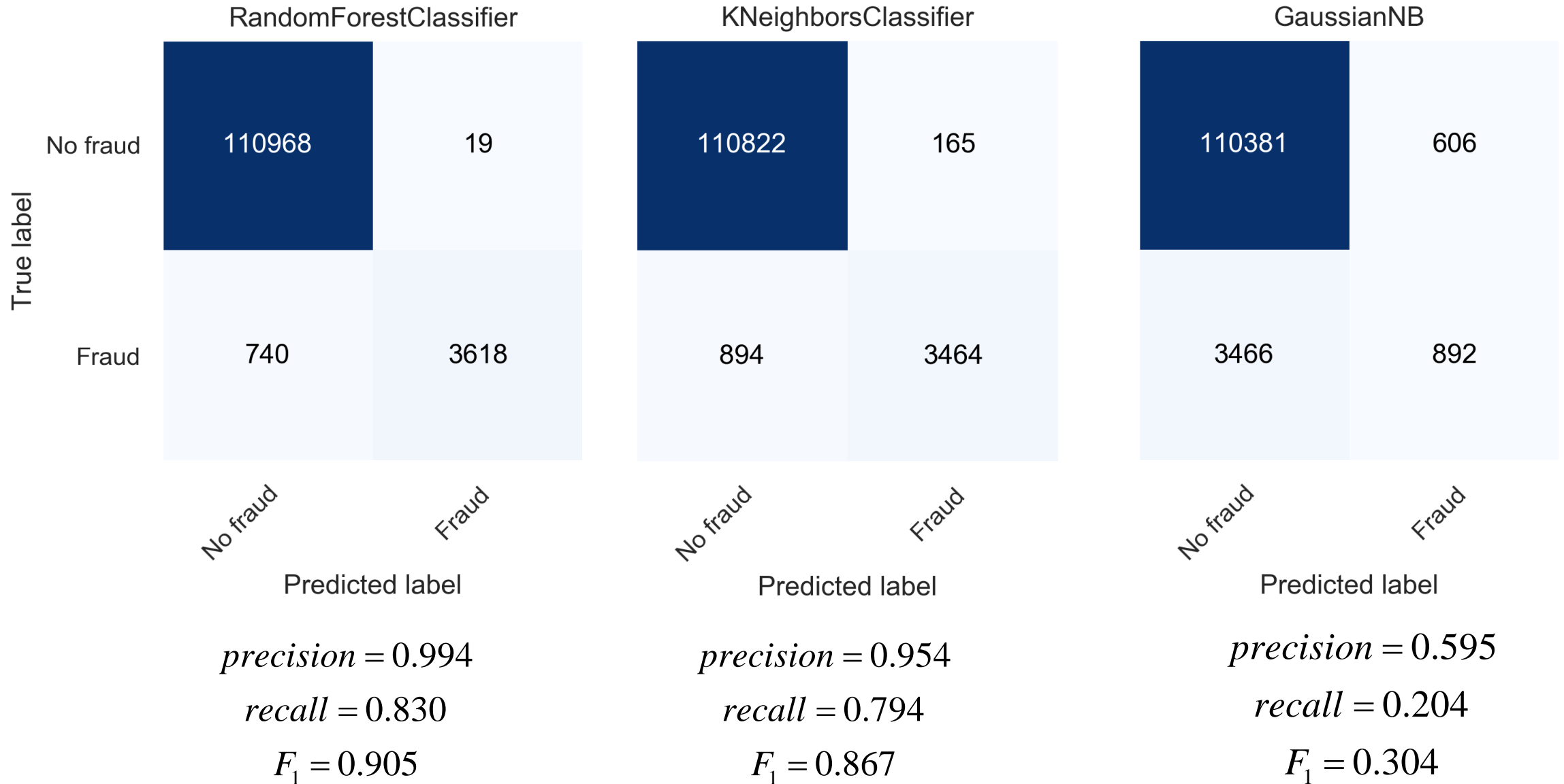
Убранные признаки: step, ID клиентов, новый баланс отправителя, старый и новый баланс получателя  
Добавленные признаки:

- **hour** шаг был конвертирован в 24 часов формат
- **newSender** первое ли появление отправителя
- **newReciver** первое ли появление получателя
- **merchant** является ли получатель магазином
- **fraudEarly** были ли раньше клиенты текущей транзакции замечены в мошеннических
- **LTS** время с момента предыдущей транзакции в качестве отправителя
- **LTR** время с момента предыдущей транзакции в качестве получателя
- **IZoB** остается ли 0 на балансе у отправителя

# Используемые методы классификации

- Случайный лес (Random forest)
- Наивный байесовский классификатор
- Метод k-ближайших соседей (K-neighbors)

# Результаты на «сырых» данных



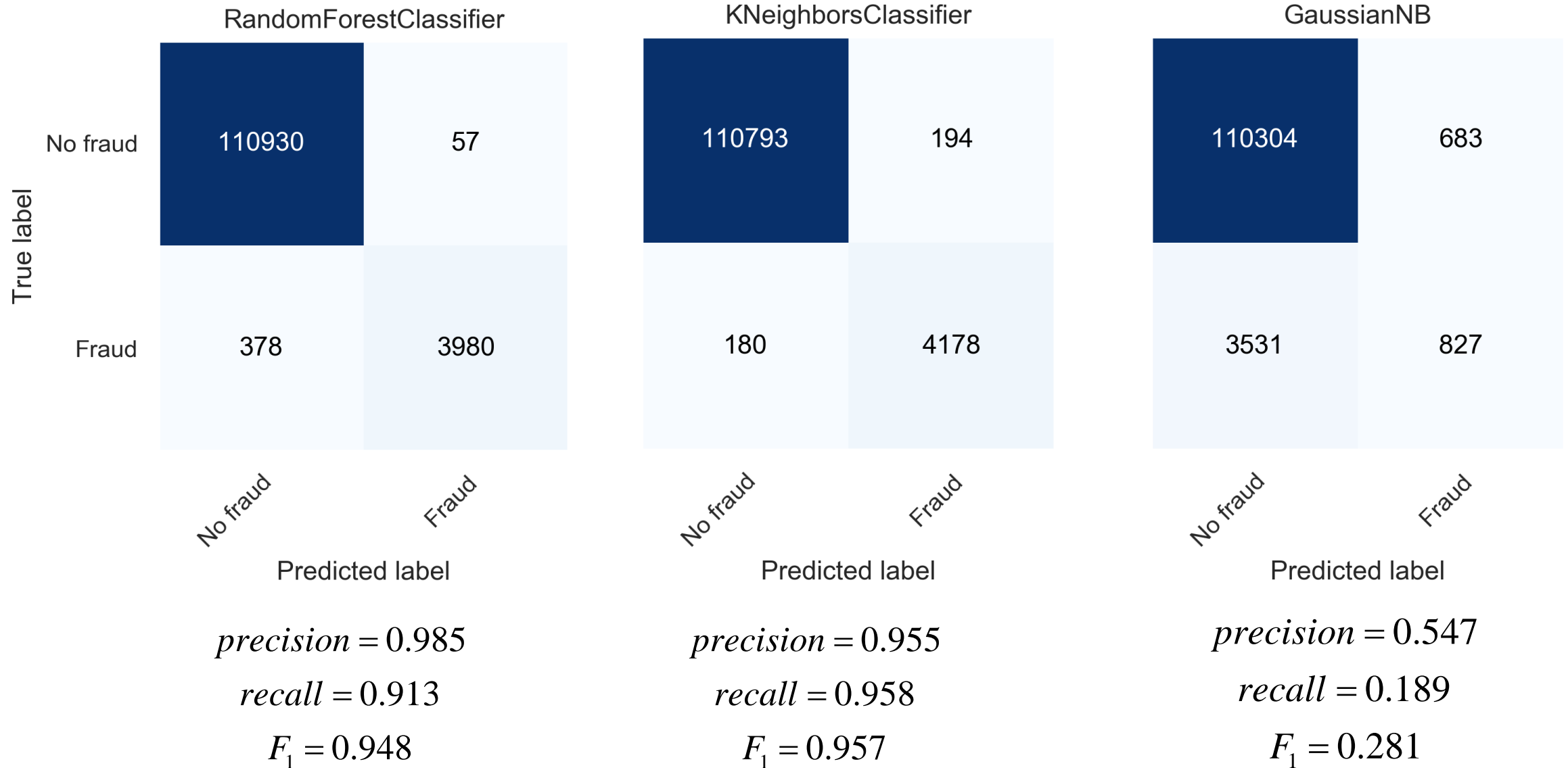


# Отбор признаков

	step	type	amount	oldbalanceOrg	isFraud	hour	newSender	newReceiver	merchant	fraudsEarly	LTS	LTR	IZoB
0	1	4	181.00	181.0	1	1	1	1	0	0	0	0	1
1	1	1	181.00	181.0	1	1	1	1	0	0	0	0	1
2	1	3	7107.77	183195.0	0	1	1	1	1	0	0	0	0
3	1	3	671.64	15123.0	0	1	1	1	1	0	0	0	0
4	1	3	1373.43	13854.0	0	1	1	1	1	0	0	0	0
Значимость признаков													
0.08*	0.06	0.24	0.38		0.11	1.25e-5	0.012	0.007	0.0004	2e-7	0.015	0.063	

RFE (recursive feature elimination) – рекурсивное отсеечение признаков

# Результат после отбора признаков



# Нормализация данных

$$x_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}$$

	type	amount	hour	iZoB
0	4.0	181.00	181.0	1.0
1	1.0	181.00	181.0	1.0
2	3.0	7107.77	183195.0	1.0
3	3.0	671.64	15123.0	1.0
4	3.0	1373.43	13854.0	1.0



	type	amount	hour	iZoB
0	0.015625	0.707010	0.707010	0.003906
1	0.003907	0.707091	0.707091	0.003907
2	0.000016	0.038770	0.999248	0.000005
3	0.000198	0.044368	0.999015	0.000066
4	0.000215	0.098652	0.995122	0.000072

# Результаты после нормализации

RandomForestClassifier

True label	Predicted label	
	No fraud	Fraud
No fraud	110987	0
Fraud	22	4336

 $\textit{precision} = 1.0$  $\textit{recall} = 0.994$  $F_1 = 0.997$ 

KNeighborsClassifier

True label	Predicted label	
	No fraud	Fraud
No fraud	110977	10
Fraud	110	4248

 $\textit{precision} = 0.997$  $\textit{recall} = 0.974$  $F_1 = 0.986$ 

GaussianNB

True label	Predicted label	
	No fraud	Fraud
No fraud	110884	103
Fraud	4333	25

 $\textit{precision} = 0.195$  $\textit{recall} = 0.005$  $F_1 = 0.011$

# Результаты

- Анализ финансовых транзакций с целью предотвращения мошеннических операций
- Прикладная программа, визуализирующая графы переводов
- Модель, способная классифицировать финансовые транзакции
- Процедуры по улучшению качества предсказаний, принесшие результат
- Сравнительный анализ эффективности разных моделей

Спасибо за внимание!