

# Warning Signals in Credit Fraud Detection

*Aneta Ptak-Chmielewska*

*Anna Matuszyk*

Edinburgh, 28-30<sup>th</sup> August 2013

# Agenda

- Purpose of the project
- Variables used
- Methods chosen
- Accuracy classification
- Fraudulent and non-fraudulent customers' profile
- Q&A

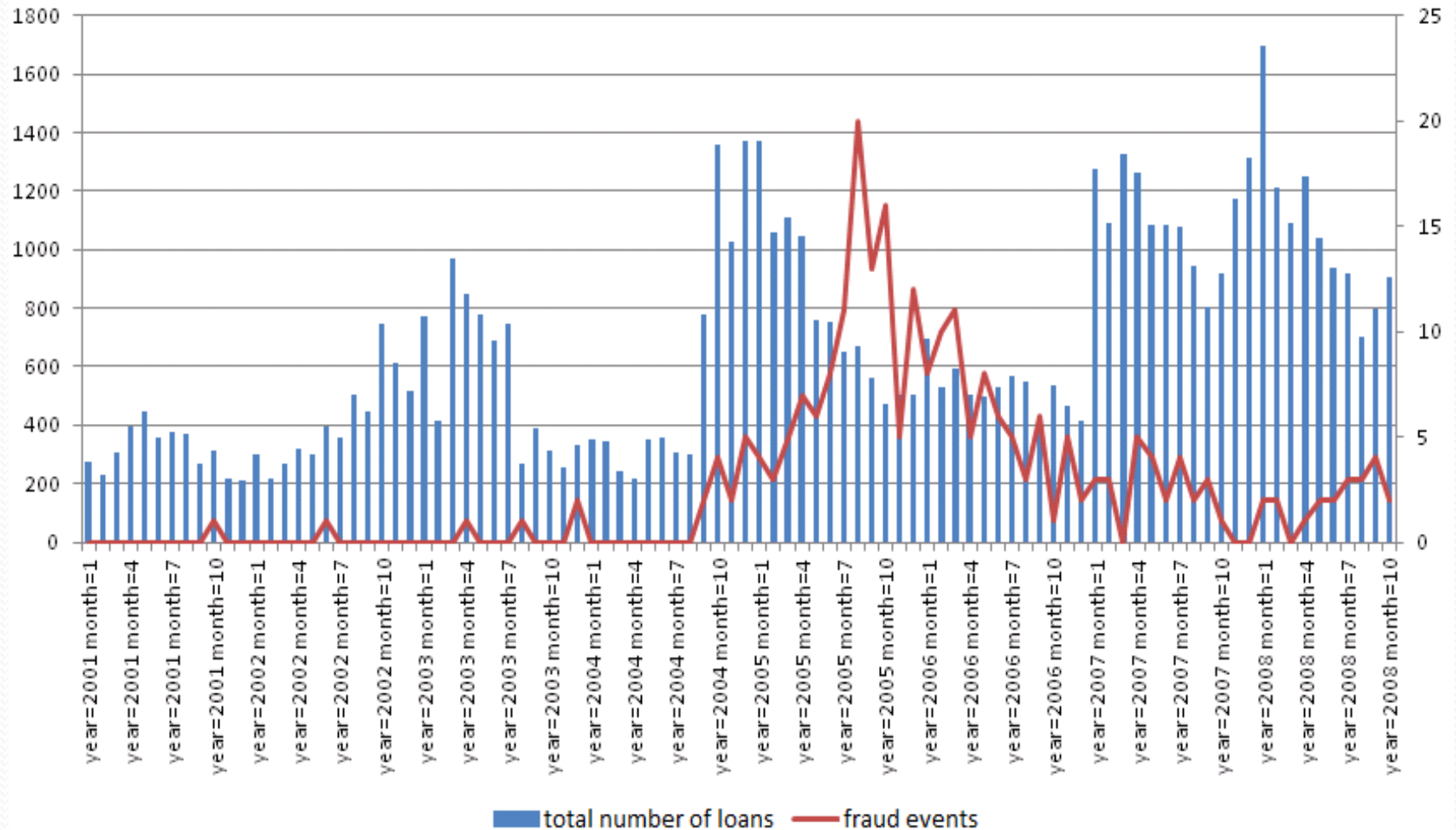
# Purpose of the Project

- Identification of potential warning signals (determinants)
- Profile of the fraudulent and non-fraudulent customer

# Data Description

- Real dataset from one European bank,
- Time range: from January 2001 to October 2008,
- ~ 65 000 cases,
- Product: automobile loans,
- 245 fraud events / 980 all cases

# Fraudulent Transactions



# Variables Used

- Brand
- Category of contract
- Gender
- Marital status
- Commercial phone number given
- No of scoring
- Children
- Type of object (used/new)
- Other securities
- Payment
- Second applicant
- Type of contract
- Customer
- Income
- Financing amount
- Duration of loan
- Purchasing price amount
- Downpayment
- Age
- Year of contract

# Methods Chosen to Build the Models

- Logistic regression (LR),
- Decision trees (DT),
- Neural network (NN).

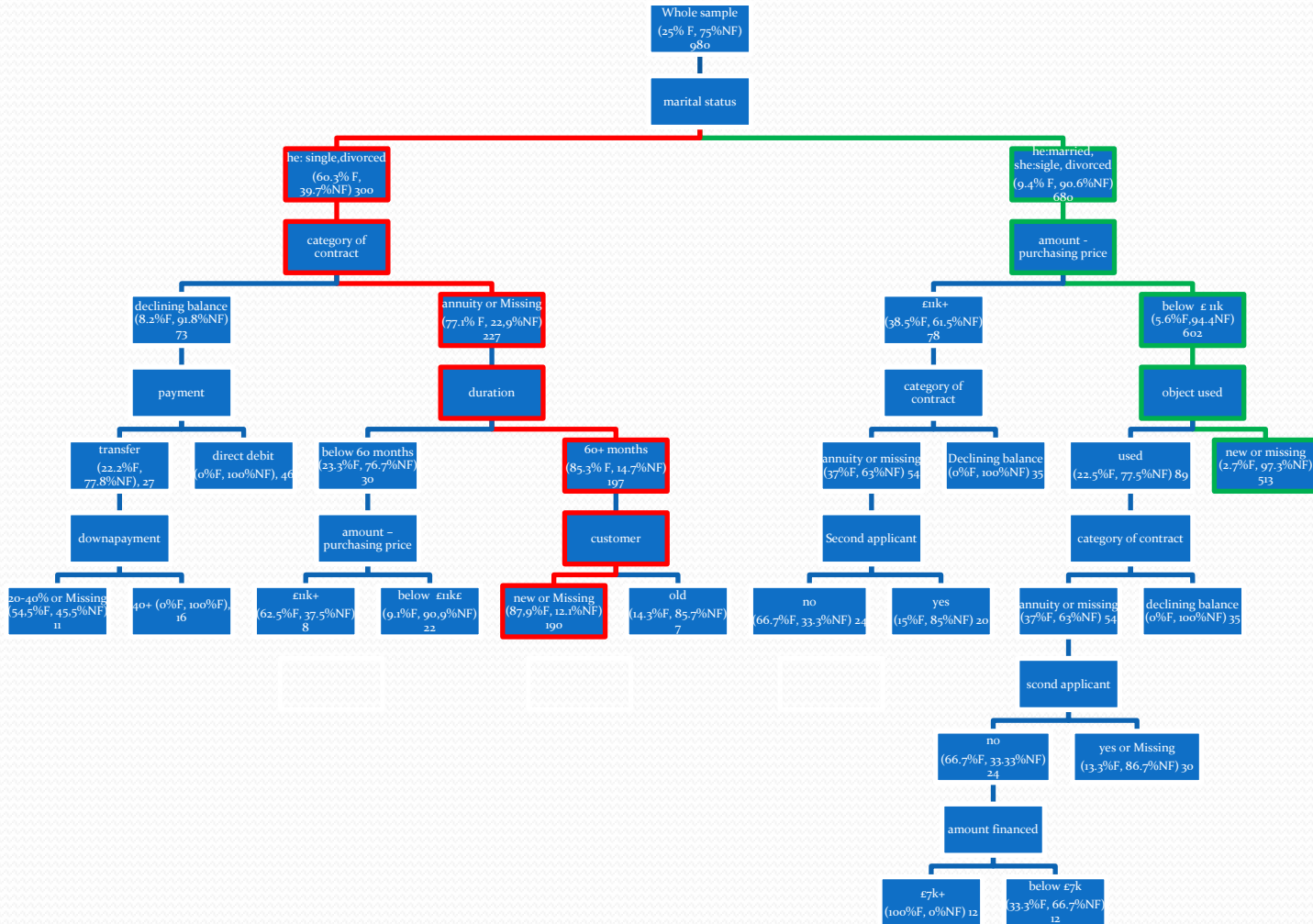
# Logistic Regression Results

Variable	Attributes	Odds ratio	p-value
<i>Describing the loan</i>			
Category of contract	Descending/no data Annuity (ref)	0.159	0.0079
Type of contract	other standard (ref)	0.318	0.0134
Payment	Direct Debit / no information transfer (ref)	0.022	0.0010
Duration of loan	below 24 months	0.089	0.3774
	<b>24-48 months</b>	<b>0.064</b>	<b>0.0588</b>
	48-60 months	0.222	0.7530
	60 months and longer (ref)		
Downpayment	below 10%	5.095	0.4049
	<b>10 -20 %</b>	<b>11.995</b>	<b>&lt;.0001</b>
	20 -40 %	3.861	0.9536
	above 40% (ref)		
Second applicant	<b>Yes</b> No (ref)	<b>0.113</b>	<b>&lt;.0001</b>

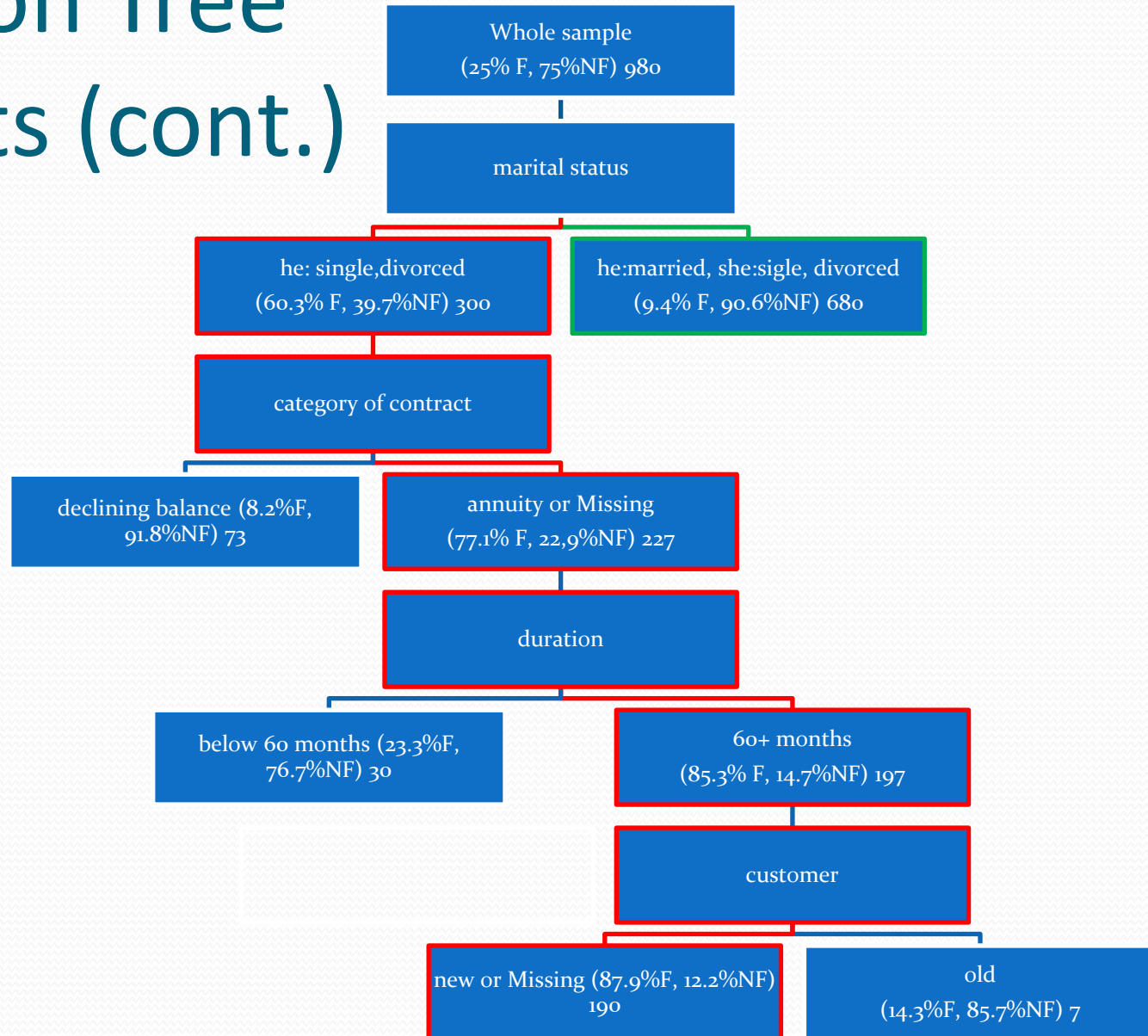
# Logistic Regression Results

Variable	Attributes	Odds ratio	p-value
<i>Describing the customer</i>			
Marital status	<b>he:single/widowed/divorced</b>	<b>2.785</b>	<b>0.0042</b>
	she:married/widowed	0.924	0.5059
	she:single/divorced	0.854	0.3494
	he:married (ref)		
Children	no data/no information	0.296	0.1006
	at least one child	0.513	0.8971
	no children (ref)		
<i>Describing the object</i>			
Type of object	<b>Used car</b>	<b>11.479</b>	<b>&lt;.0001</b>
	New car (ref)		
Purchasing price amount	<b>11 k£ +</b>	<b>4.993</b>	<b>&lt;.0001</b>
	7 k£-11 k£	1.394	0.1434
	below 7 k£ (ref)		

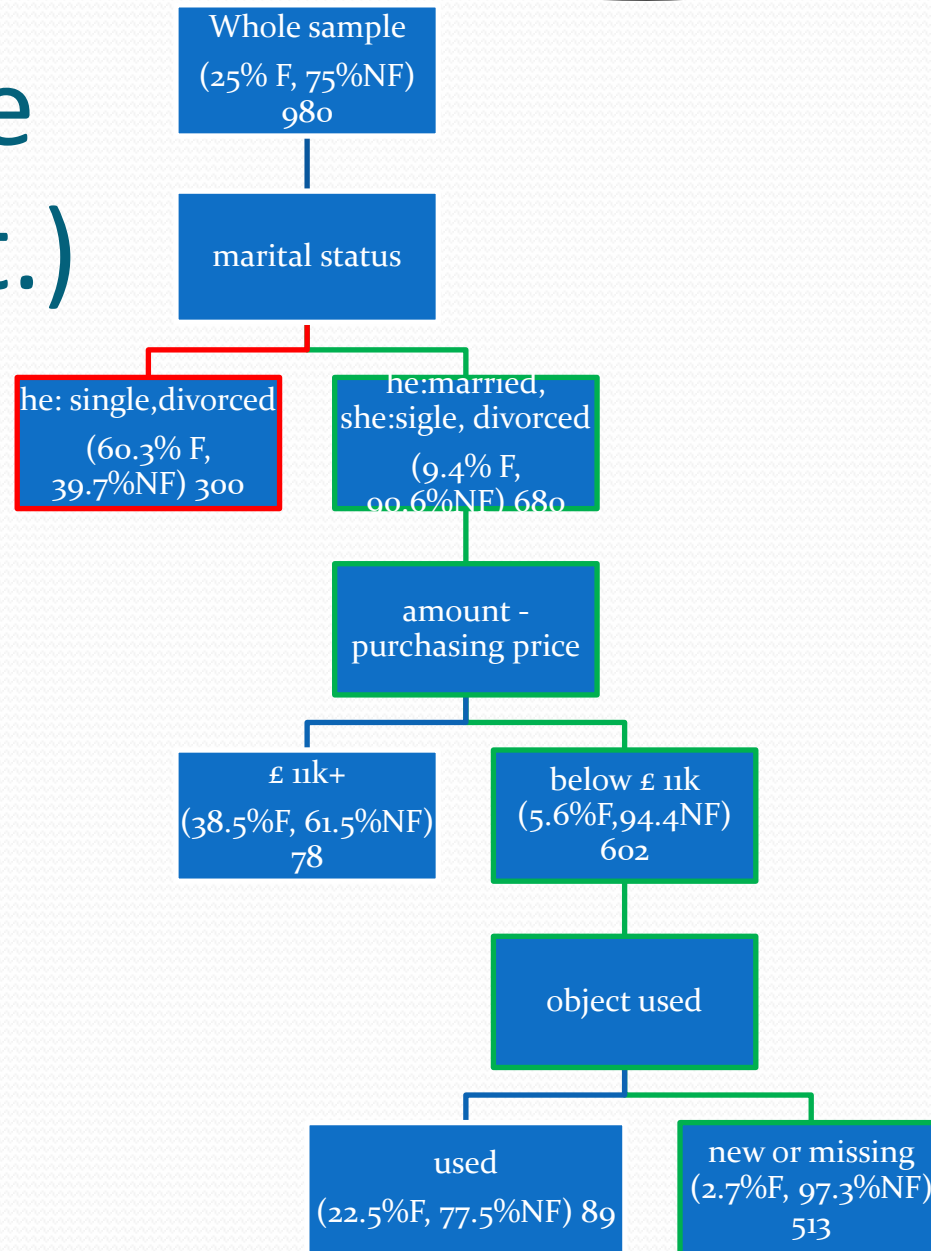
# Decision Tree Results



# Decision Tree Results (cont.)



# Decision Tree Results (cont.)



# Selection of the Model

- Significant variables in the DT model confirmed results accuracy obtained in the LR model,
- Difficult parameters interpretation in the NN model,
- Quality of the NN model similar to the quality of the LR model,
- LR and DT models usage

# Fraudulent Customer Profile:

- he: single/widowed/divorced,
- category of contract: annuity,
- duration: 60+ months,
- new customer,
- purchasing price: £ 11k+,

190/980 cases (19.4%), (whole sample 25% ), probability of fraud: 87.9% ; 3.5 times higher the chance of fraud in comparison with the whole sample,

# Non-fraudulent Customer Profile:

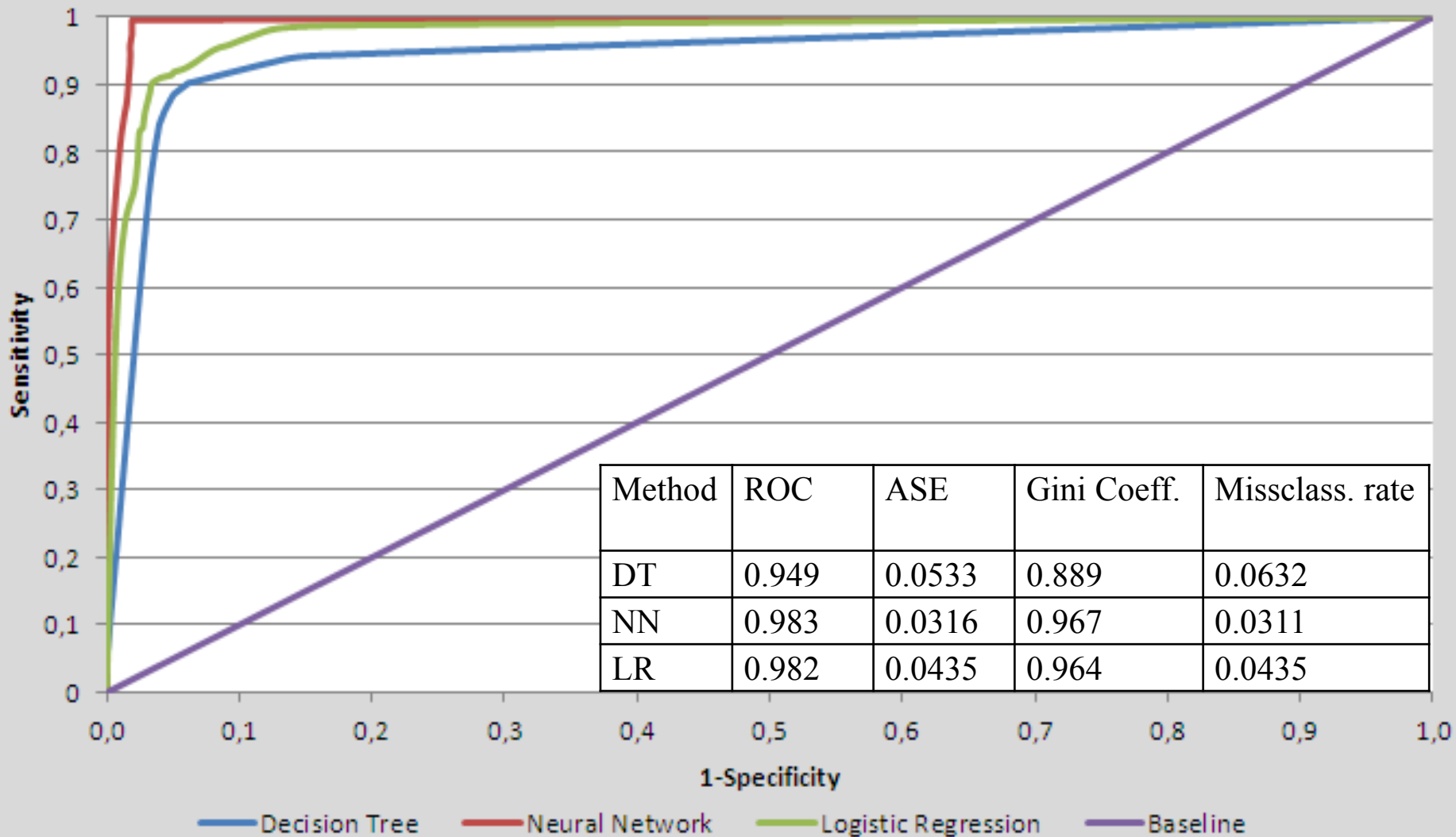
- she: married/widowed/single/divorced or he: married,
- purchasing price < £ 11k,
- new car,

513 / 980 cases (52.3%); probability of fraud: 2.7%; almost 10 times lower the chance of fraud in comparison with the whole sample (25%)

# Accuracy Classification

Method Used	Actual G/ Predicted G	Actual G/ Predicted F	Actual F/ Predicted G	Actual F/ Predicted F
Actual	735	-	-	245
DT	698	37	28	217
NN	721	14	2	243
LR	710	25	24	221

# Performance Measures



# Problems with Modelling Application Fraud

- Very limited literature
- Difficulty in obtaining the data from the banks (confidentiality reasons)
- Risk, that fraudsters use the research results and change their behaviour
- Large data sets with tiny proportion of fraudulent transactions

# Future Research

- Development process continuation
- Adding additional data
- Including macro-economic variables
- Using new statistical techniques

# Q & A



Contacts:

[aptak@sgh.waw.pl](mailto:aptak@sgh.waw.pl)

[anna.matuszyk@sgh.waw.pl](mailto:anna.matuszyk@sgh.waw.pl)