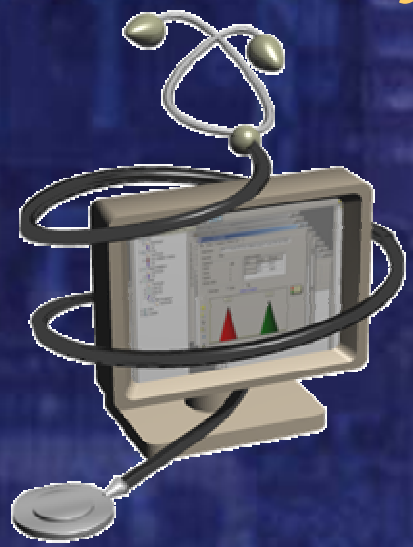
FUJITSU

Ghost Miner Software

GhostMiner's application in life sciences



FQS Poland

FUJITSU

Ghost Miner Software

Example: Gestational Diabetes Mellitus prediction

FQS Poland

FUJITSU

Ghost Miner Software

Gestational Diabetes Mellitus

background

- *third type of diabetes mellitus*
- *defined as any carbohydrates intolerance recognized first time during pregnancy*
- *occurs in about 3% to 12% of pregnancies*
- *women with gestational diabetes have significantly different pregnancy characteristics and outcomes: intrauterine fetal death, fetal macrosomia and growth disorder, neonatal hypoglycemia ie.*

FQS Poland

Gestational Diabetes Mellitus

risk factors

- *obesity or overweight*
- *maternal age*
- *history of miscarriages*
- *life style*
- *family history of diabetes*

FUJITSU

Ghost Miner Software

FQS Poland

FUJITSU

Ghost Miner Software

Gestational Diabetes Mellitus

prediction - today

- *regional experts and scientific organizations recommend **screening** as a good clinical practice. For example Experts Committee of Polish Diabetes Association recommends fasting glucose blood concentration measurements in all pregnant women during first visit in consulting room*
- *disadvantages – price and onerousness as well as for women and health specialists*

FQS Poland

FUJITSU

Ghost Miner Software

Gestational Diabetes Mellitus

prediction – tomorrow?

- *tools: computer-based, intelligent, problem-specific, able to learn and available both for patients and healt specialists*
- *advantages – easy to use and cost-effective methods*

FQS Poland

Gestational Diabetes Mellitus

prediction – tomorrow?

GhostMiner™

FQS Poland

FUJITSU

Ghost Miner Software

Gestational Diabetes Mellitus

project aim

to use novel modeling techniques as flexible and advanced methods helpful in GDM prediction based on simple demographic information

FQS Poland

Gestational Diabetes Mellitus

data

- *epidemiological databases obtained from National Center for Health Statistics (NCHS)*
- *main database - BRFSS database (Behavioral Risk Factor Surveillance System)*
- *BRFSS is an ongoing data collection program designed to measure behavioral risk factors in the adult population (18 years of age or over) living in households*
- *GDM prediction inclusion criteria - at least one pregnancy in last 5 years, no other type of diabetes diagnosed*

FUJITSU

Ghost Miner Software

FQS Poland

FUJITSU

Ghost Miner Software

Gestational Diabetes Mellitus

data

- *2551 records*
- *91 women were diagnosed with gestational diabetes*
- *selection of crucial factors governing analyzed problem based on an expert opinion and professional literature - **8 variables** from ca. 200 were selected to the investigation*
- *label 1 denoted diagnosed diabetes and label 0 lack of diabetes*

FQS Poland

FUJITSU

Ghost Miner Software

Gestational Diabetes Mellitus

data preprocessing

due to specific structure of native dataset (unequal distribution of cases, it was necessary to alter learning datasets; two different ways of datasets modification were tested:

- ***„balance minus"*** *– randomly chosen records from more frequently represented class were removed from learning dataset*
- ***„balance plus"*** *- randomly chosen records from less frequently represented class from learning dataset were duplicated*

FQS Poland

No of variable	Description	All patients	GDM positive	GDM negative
1	**High blood pressure**	2551	91	2460
	YES	298	17 (6%)	281 (94%)
	NO	2253	74 (3%)	2179 (97%)
2	**High cholesterol (YES/NO)**			
	YES	223	8 (4%)	215 (96%)
	NO	1281	39 (4%)	1242 (96%)
	REFUSED	1047	44 (4%)	1003 (96%)
3	**Do you now smoke cigarettes**			
	EVERYDAY	575	12 (2%)	563 (98%)
	SOME DAYS	118	7 (6%)	111 (94%)
	NOT AT ALL	321	14 (4%)	307 (96%)
	REFUSED	1537	58 (4%)	1479 (96%)
4	**Age in years (18 - 44)**	30,00 (SD = 6,01)	34,11 (SD = 5,60)	30,00 (SD = 6,02)
5	**Race**			
	CAUCASIAN	2023	76 (4%)	1947 (96%)
	BLACK	396	12 (3%)	384 (97%)
	ASIAN	32	2 (6%)	30 (94%)
	AMERICAN INDIAN	35	0 (0%)	35 (100%)
	OTHER	65	1 (2%)	64 (98%)
6	**Weight [kg]**	67,69 (SD = 15,22)	69,34 (SD = 14,99)	67,63 (SD = 15,23)
7	**Height [cm]**	165,00 (SD = 6,89)	163,70 (SD = 7,69)	165,04 (SD = 6,86)
8	**Are you eating fewer fat to lower your risk of developing heart disease or stroke**			
	YES	867	30 (3%)	837 (97%)
	NO	468	17 (4%)	451(96%)
	REFUSED	1216	44 (4%)	1172 (96%)

FUJITSU

Ghost Miner Software

FQS Poland

FUJITSU

Ghost Miner Software

Gestational Diabetes Mellitus

modeling methods

- *Support Vector Machine (SVM)*
- *Separability of Split Value (SSV) decision trees*
- *k-nearest-neighbors algorithm (kNN)*
- *ANN algorithms (IncNET, FSM)*

FQS Poland

Gestational Diabetes Mellitus

results

- *„global" - percentage of all positive model answers*
- *„0" - percentages of correctly classified patients without*
- *„1" - percentages of correctly classified patients with GDM*
- *10-fold cross validation technique was applied*

Gestational Diabetes Mellitus

results

Nr	Modeling method	Dataset	Generalization test results		
			0	1	Global
1	SSV Tree	bal-	**59.6**	**55.6**	**57.8**
2	FSM	bal-	**47.85**	**49.9**	**48.1**
3	kNN	bal-	76.1	27.8	52.3
4	IncNET	bal-	**74.6**	**36.65**	**56.4**
5	kNN	bal+	12.50	96.0	94.0
6	IncNET	bal+	37.50	58.0	57.0
7	SVM	bal+	12.50	97.0	94.0

FUJITSU

Ghost Miner Software

FQS Poland

FUJITSU

Ghost Miner Software

Gestational Diabetes Mellitus

*comparison with **logistic regression***

- *classical modeling method when binary outcome is observed*
- *in this study logistic regression analysis was used to predict incidence of GDM on the basis of 8 chosen inputs and „balance plus" data sets due to better results than other modeling methods*
- *the dependent variable was the dichotomized GDM diagnosis value*
- *the independent variables entered into the analysis were the same as for other modeling techniques*

FQS Poland

Gestational Diabetes Mellitus

comparison with logistic regression

Test file	Global	0	1	Threshold
1	59.19	59.54	50.00	
2	59.93	60.15	54.55	
3	56.62	56.76	53.85	
4	59.19	59.54	50.00	
5	54.41	54.75	50.00	
6	56.99	57.03	62.50	0.5
7	51.84	52.26	40.00	
8	57.35	57.98	50.00	
9	57.35	57.92	50.00	
10	51.47	51.34	60.00	
Mean	**56.43**	**56.73**	**52.09**	

FUJITSU

Ghost Miner Software

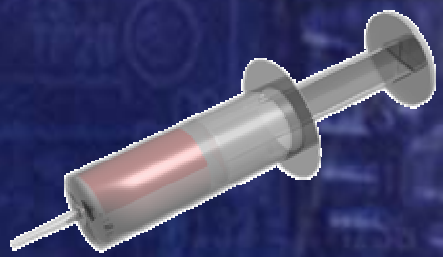
FQS Poland

FUJITSU

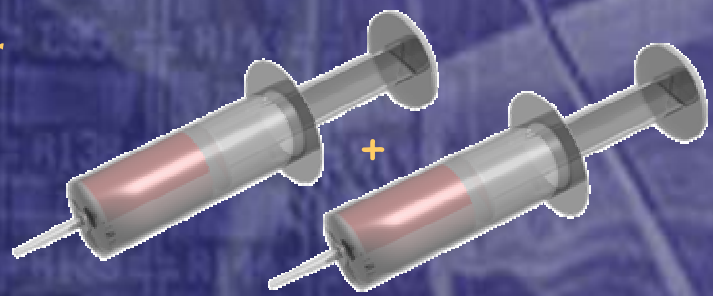
practical application

prediction - today

PREGNANT WOMAN



or



or even

Ghost Miner Software

FQS Poland

FUJITSU

Ghost Miner Software

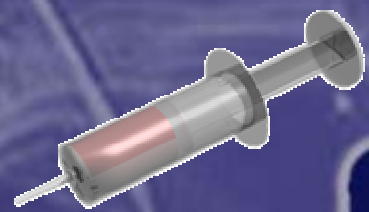
practical application

prediction – tomorrow

Internet or locally available

PREGNANT WOMAN

if positive



BASED ON INTELLIGENT TECHNIQUES GDM-SCREENING SYSTEM

FQS Poland

Gestational Diabetes Mellitus

conclusions

- ***GhostMiner*** *software is a good source of novel, easy to use and powerful data-analysis tools*
- *owing to advanced model betterment abilites implemented in* ***GhostMiner*** *software – obtained results are not definitive*
- *further investigations in health science area with intelligent prediction techniques are necessary*

FUJITSU

Ghost Miner Software

FQS Poland

FUJITSU

Ghost Miner Software

Any questions left?

Ask us!

gm@fqspl.com.pl

Or visit our webpage

http://www.fqspl.com.pl/GhostMiner

FQS Poland