



Politechnika Wroclawska

A review of process metrics in defect prediction studies

Przegląd zastosowań metryk procesu w badaniach dotyczących predykcji defektów

Marian Jureczko
Lech Madeyski

Politechnika Wroclawska

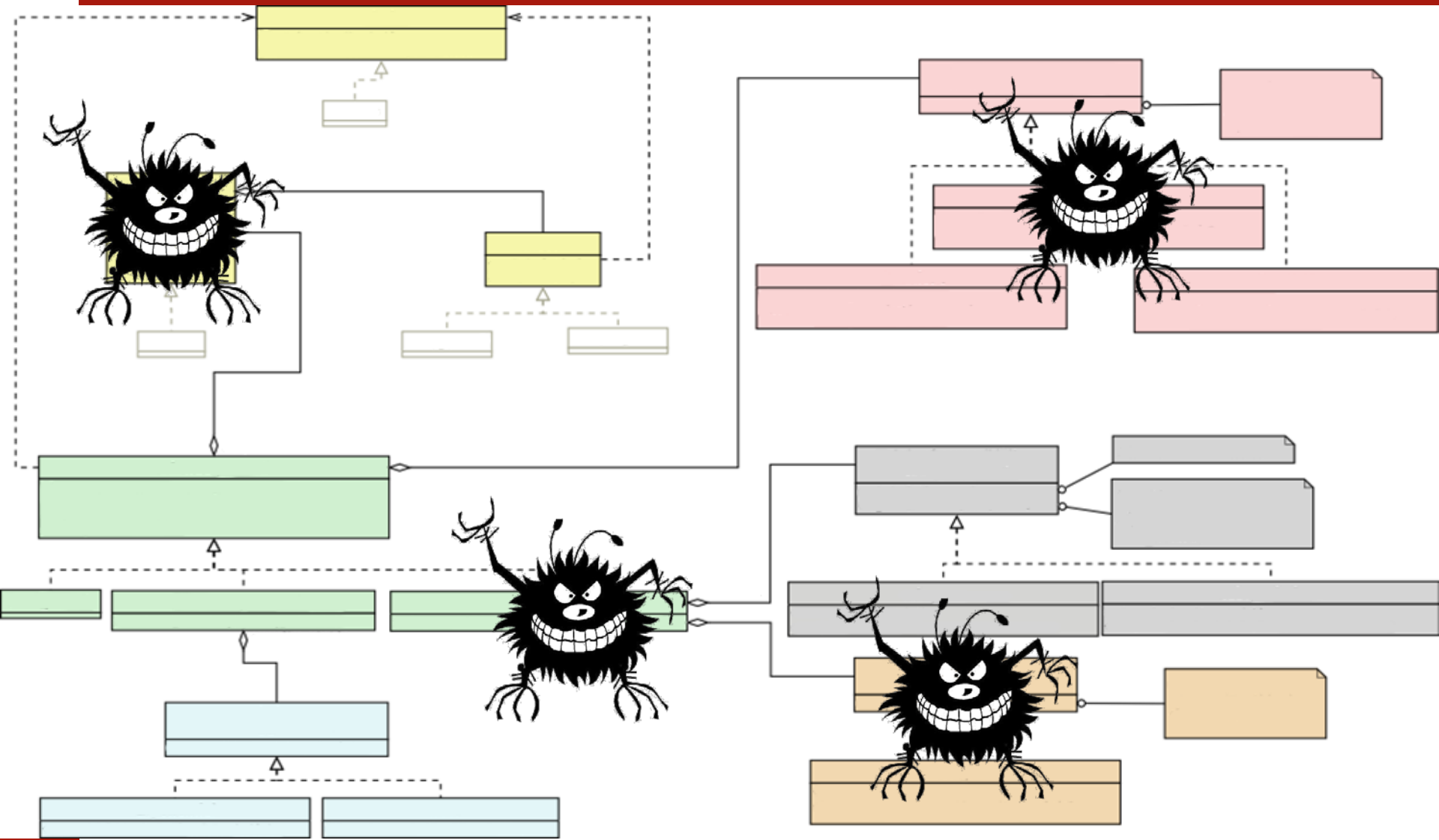


Agenda

- Wprowadzenie do modeli predykcji defektów
- Definicje metryk
- Metryki procesu w modelach predykcji defektów
- Statystyki opisowe
- Podsumowanie

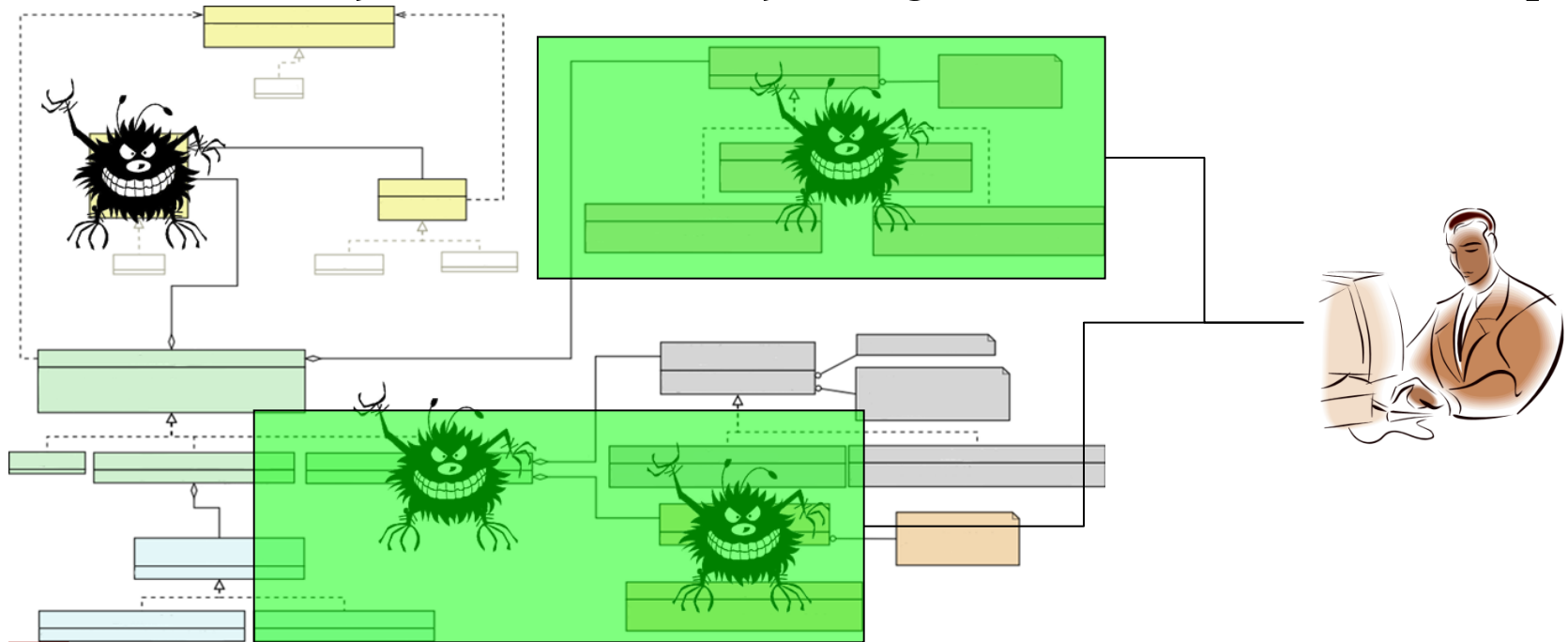


Wprowadzenie



Motywacja - Dlaczego predykcja defektów?

20% modułów systemu informatycznego zawiera 80% defektów[1]



Można użyć modelu predykcji defektów do wyszukiwania klas obciążonych defektami i tym samym zoptymalizować testy.

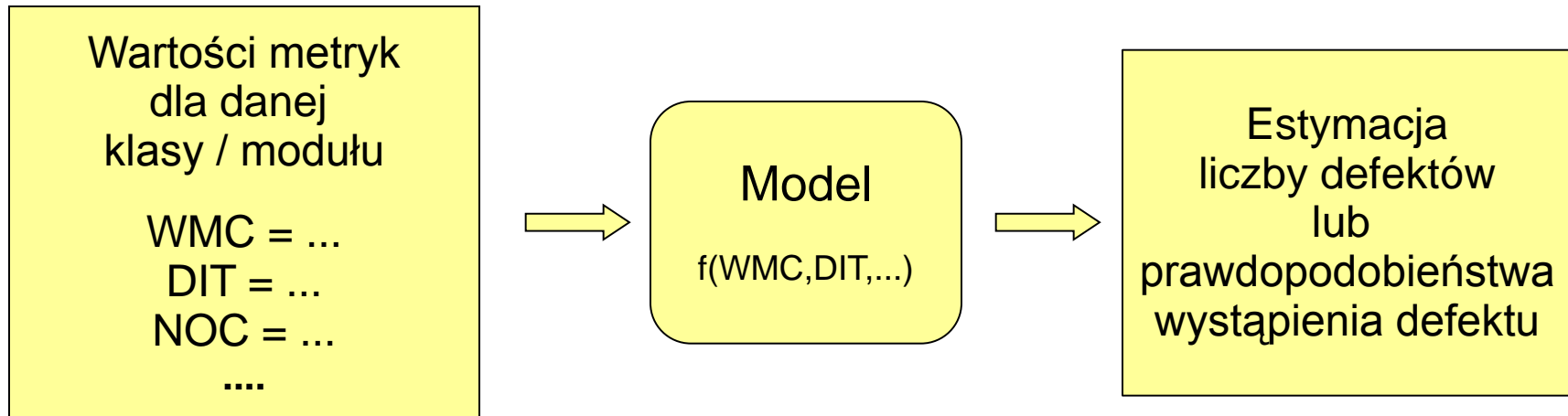


Definicje

•Defekt

- Defekt to anomalia w produkcie, która może być brakiem znalezionym we wczesnych etapach prac lub błędem w funkcjonowaniu produktu[1]
- Defekt to zagadnienie, które zostało zaraportowane i oznaczone jako błąd w systemie śledzenia defektów używanym w projekcie, w którym zagadnienie zostało zidentyfikowane[2]

•Model predykcji defektów



[1]982.2 IEEE/ANSII

[2]Nikora Munson, *The effects of fault counting methods on fault model quality*, COMPSAC '04



Agenda

- Wprowadzenie do modeli predykcji defektów
- Definicje metryk
- Metryki procesu w modelach predykcji defektów
- Statystyki opisowe
- Podsumowanie



Metryki procesu

- Liczba historycznych wersji, ang. Number of Revisions (NR)
- Liczba różnych autorów, ang. Number of Distinct Committers (NDC)
- Liczba zmodyfikowanych linii, ang. Number of Modified Lines (NML)
- Czy klasa jest nowa, ang. Is New (IN)
- Liczba defektów naprawionych w poprzednim wydaniu projektu, ang. Number of Defects fixed in Previous Version (NDPV)



Taksonomia metryk procesu

Metryka	Kod źródłowy	Programista	Defekt
NR	X		
NML	X		
IN	X		
NDC	X	X	
NDPV	X		X



Agenda

- Wprowadzenie do modeli predykcji defektów
- Definicje metryk
- Metryki procesu w modelach predykcji defektów
- Statystyki opisowe
- Podsumowanie



Liczba historycznych wersji (NR)

Artykuł	Zbiór danych (Przem. Otw.)		Cel badań	Rekomendacja
Bel et al. [3]	1	0	Predykcja, korelacja	↗
Ostrand et al. [4]	1	0	Predykcja	-
Weyuker et al. [5]	1	0	Predykcja	↗
Weyuker et al. [6]	1	0	Predykcja	-
Weyuker et al. [7]	3	0	Predykcja	-
Ratzinger et al. [8]	1	0	Predykcja	-
Graves et al. [9]	1	0	Predykcja	↗
Schroeter et al. [10]	0	1	Korelacja	-
Illes-Seifert and Paech [2]	0	9	Korelacja	↗
Shihab et al. [13]	0	1	Predykcja	↗
Moser et al. [14]	0	1	Predykcja	↗
Kim et al. [34]	0	7	Predykcja	-
Gao et al. [36]	1	0	Selekcja cech	↗
Jureczko [37]	7	15	Predykcja, korelacja	↗



Liczba zmodyfikowanych linii (NML)

Artykuł	Zbiór danych (Przem. Otw.)		Cel badań	Rekomendacja
Khoshgoftaar et al. [19]	2	0	Predykcja	↗
Layman et al. [20]	1	0	Predykcja	↗
Nagappan and Ball [21]	1	0	Predykcja	↗
Nagappan et al. [18]	1	0	Predykcja	-
Nagappan et al. [12]	1	0	Predykcja	↗
Purushothaman et al. [22]	1	0	Rozkład błędów	↗
Zimmermann et al. [15]	7	4	Predykcja	-
Ratzinger et al. [8]	1	0	Predykcja	-
Śliwerski et al. [23]	0	2	Predykcja	↗
Hassan [24]	0	6	Predykcja	↗
Giger et al. [25]	0	15	Predykcja	↗
Giger et al. [28]	0	16	Predykcja	-
Graves et al. [9]	1	0	Predykcja	↗
Moser et al. [14]	0	1	Predykcja	↗



Liczba zmodyfikowanych linii (NML)

Artykuł	Zbiór danych (Przem. Otw.)		Cel badań	Rekomendacja
Bell et al. [39]	1	0	Predykcja	↗
Shin et al. [40]	0	2	Predykcja	↗
Sisman et al. [41]	0	1	Predykcja	↗
Krishnan et al. [42]	0	1	Predykcja	↗
Khoshgoftaar et al. [30]	1	0	Predykcja	↗
Gao et al. [36]	1	0	Selekcja cech	↗
Jureczko [37]	7	15	Predykcja, korelacja	↗



Czy klasa jest nowa (IN)

Artykuł	Zbiór danych (Przem. Otw.)		Cel badań	Rekomendacja
Ostrand et al. [31]	1	0	Rozkład błędów	-
Ostrand et al. [29]	2	0	Predykcja	↗
Illes-Seifert and Paech [2]	0	9	Korelacja	-
Bel et al. [3]	1	0	Predykcja, korelacja	-
Weyuker et al. [5]	1	0	Predykcja	↗
Weyuker et al. [6]	1	0	Predykcja	-
Weyuker et al. [7]	3	0	Predykcja	-
Graves et al. [9]	1	0	Predykcja	↗
Khoshgoftaar et al. [30]	1	0	Predykcja	-
Ostrand et al. [4]	1	0	Predykcja	-
Moser et al. [14]	0	1	Predykcja	↗
Kim et al. [34]	0	7	Predykcja	-



Liczba różnych autorów (NDC)

Artykuł	Zbiór danych (Przem. Otw.)		Cel badań	Rekomendacja
Weyuker et al. [6]	1	0	Predykcja	↗
Weyuker et al. [7]	3	0	Predykcja	↗
Ratzinger et al. [8]	1	0	Predykcja	-
Zimmermann et al. [15]	7	4	Predykcja	-
Graves et al. [9]	1	0	Predykcja	↘
Schroeter et al. [10]	0	1	Korelacja	-
Matsumoto et al. [16]	0	1	Efekt programisty	↗
Weyuker et al. [17]	1	0	Predykcja	-
Nagappan et al. [18]	1	0	Predykcja	↗
Illes-Seifert and Paech [2]	0	9	Korelacja	↗
Moser et al. [14]	0	1	Predykcja	↗
Bell et al. [38]	3	0	Predykcja	-
Jureczko [37]	7	15	Predykcja, korelacja	↗



Liczba defektów naprawionych w poprzednim wydaniu projektu (NDPV)

Artykuł	Zbiór danych (Przem. Otw.)		Cel badań	Rekomendacja
Wahyudin et al. [32]	0	1	Predykcja	↗
Arisholm and Briand [33]	1	0	Predykcja	↗
Ostrand et al. [31]	1	0	Predykcja	↗
Ostrand et al. [29]	2	0	Predykcja	-
Weyuker et al. [5]	1	0	Predykcja	↗
Weyuker et al. [7]	3	0	Predykcja	-
Graves et al. [9]	1	0	Predykcja	-
Kim et al. [34]	0	7	Predykcja	↗
Khoshgoftaar et al. [30]	1	0	Predykcja	↗
Gyimothy et al. [35]	0	1	Korelacja	↗
Illes-Seifert and Paech [2]	0	9	Korelacja	-
Schroeter et al. [10]	1	0	Korelacja	-



Liczba defektów naprawionych w poprzednim wydaniu projektu (NDPV)

Artykuł	Zbiór danych (Przem. Otw.)		Cel badań	Rekomendacja
Shihab et al. [13]	0	1	Predykcja	↗
Ostrand et al. [4]	1	0	Predykcja	-
Moser et al. [14]	0	1	Predykcja	↗
Sisman et al. [41]	0	1	Predykcja	↗
Gao et al. [36]	1	0	Selekcja cech	↗
Jureczko [37]	7	15	Predykcja, korelacja	↗



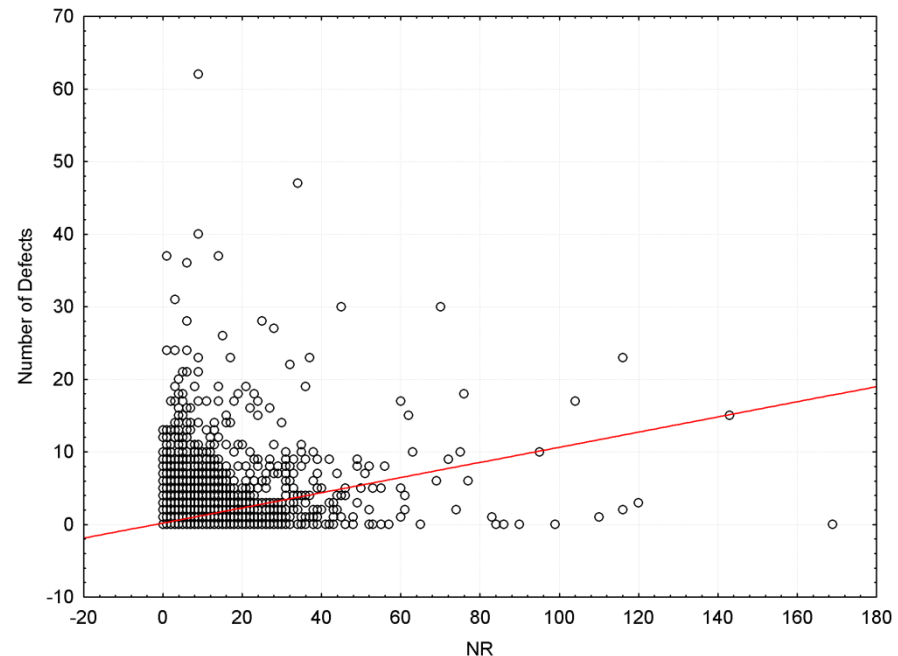
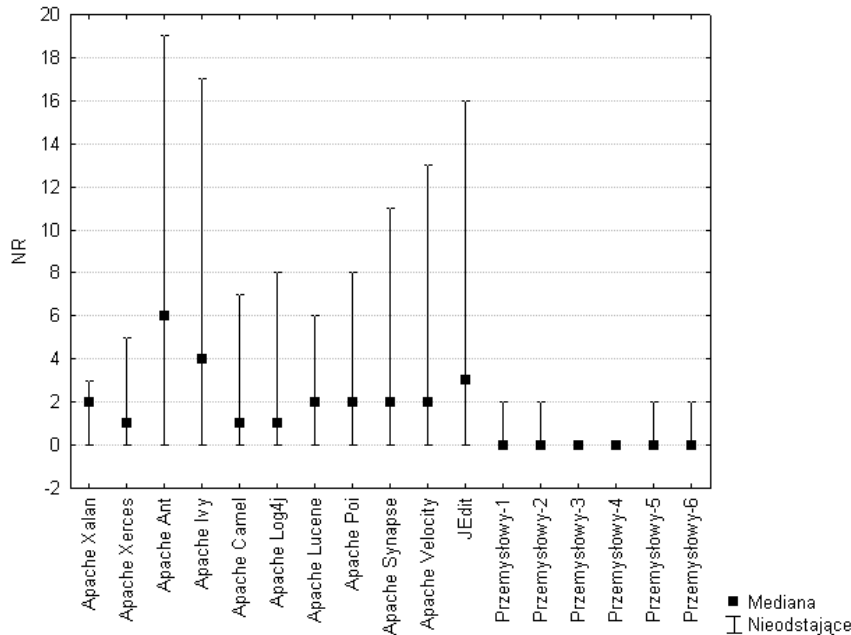
Agenda

- Wprowadzenie do modeli predykcji defektów
- Definicje metryk
- Metryki procesu w modelach predykcji defektów
- Statystyki opisowe
- Podsumowanie



Liczba historycznych wersji (NR)

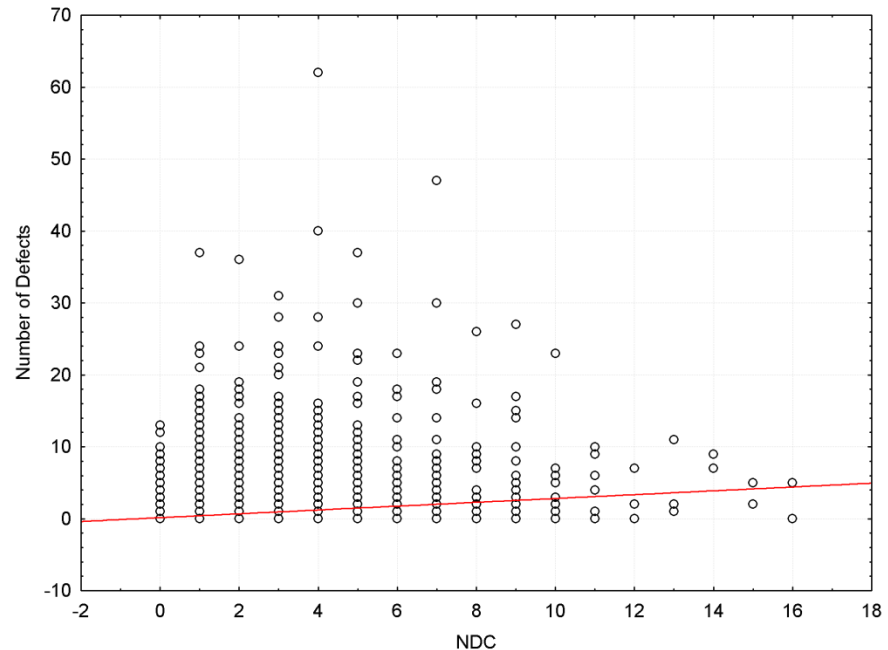
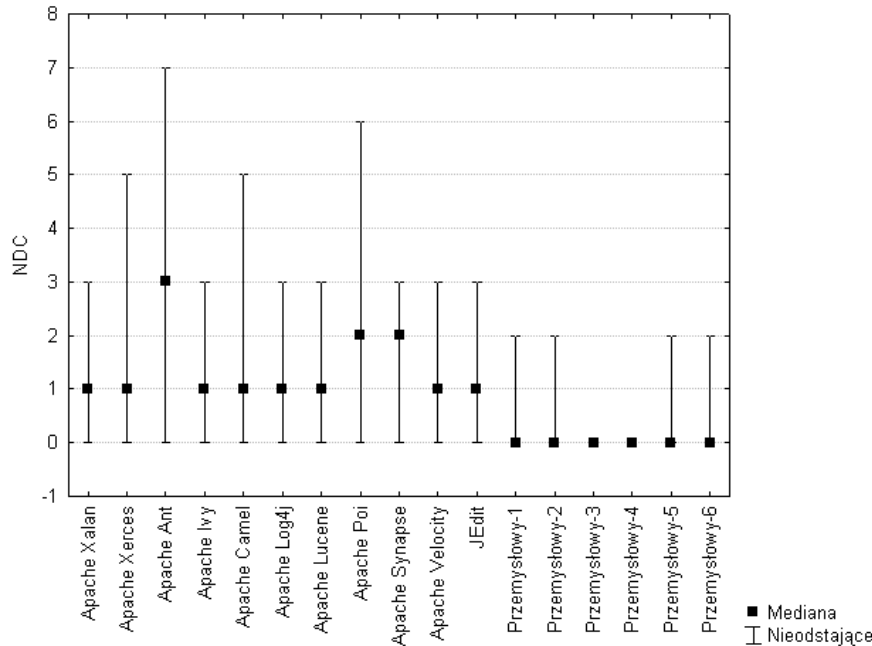
Średnia	Odchylenie standardowe	Korelacja Pearsona z liczbą defektów	Korelacja Spearmana z liczbą defektów	Liczba przebadanych projektów
1,27	3,54	0,31	0,28	17





Liczba różnych autorów (NDC)

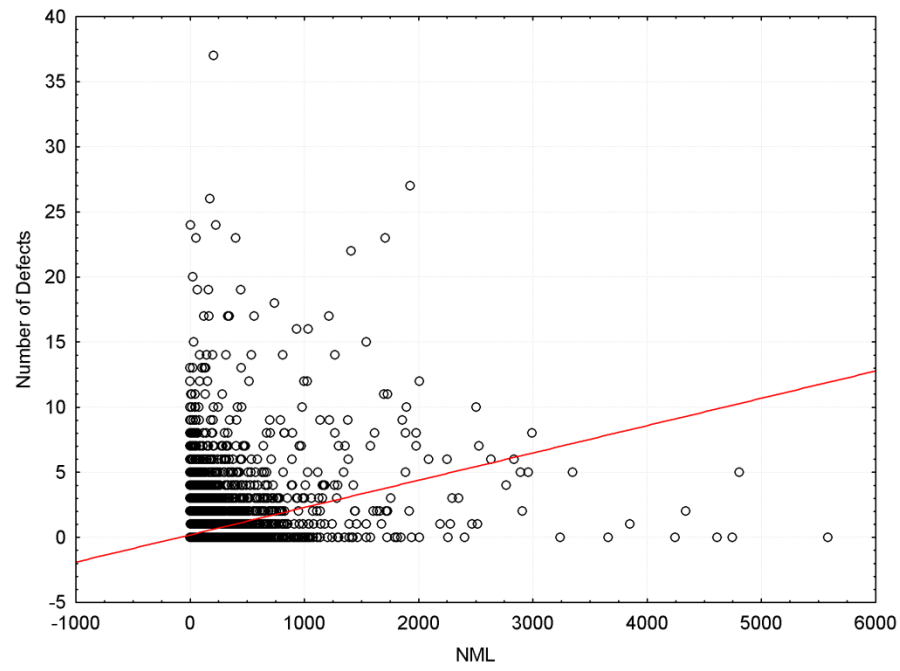
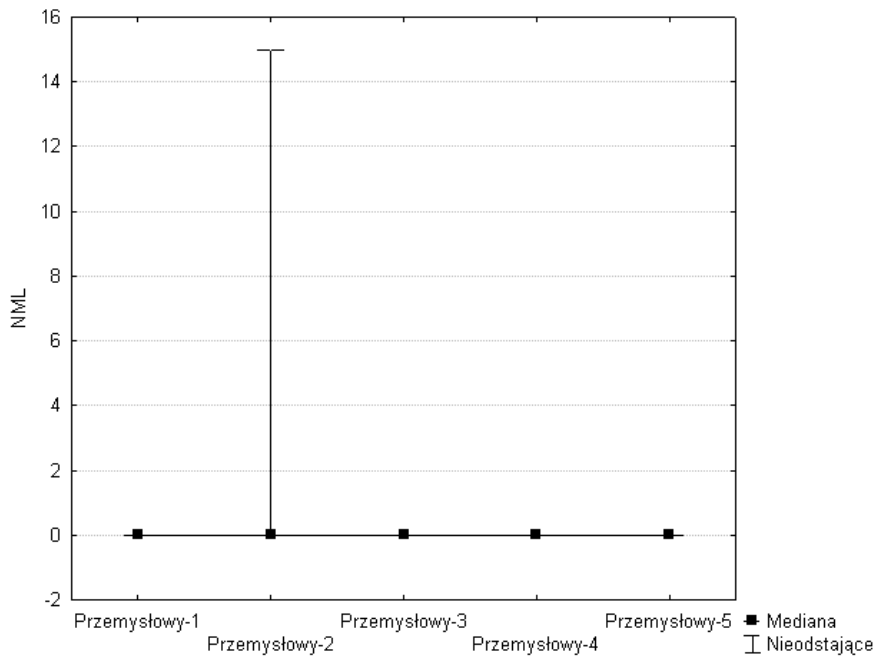
Średnia	Odchylenie standardowe	Korelacja Pearsona z liczbą defektów	Korelacja Spearmana z liczbą defektów	Liczba przebadanych projektów
0,68	1,19	0,27	0,27	17





Liczba zmodyfikowanych linii (NML)

Średnia	Odchylenie standardowe	Korelacja Pearsona z liczbą defektów	Korelacja Spearmana z liczbą defektów	Liczba przebadanych projektów
19,77	113,65	0,27	0,20	5





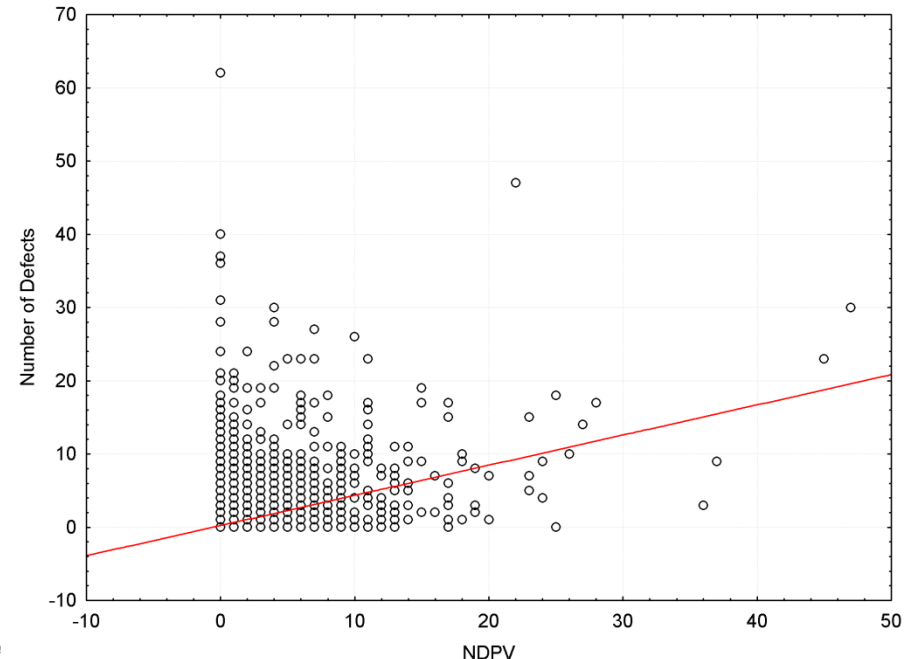
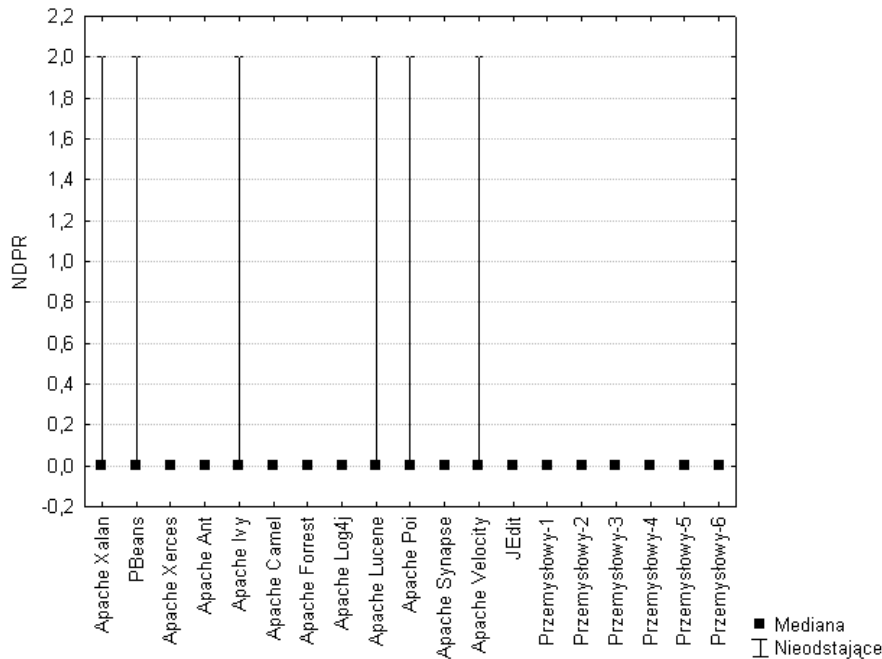
Czy klasa jest nowa (IN)

Liczba starych klas	Liczba nowych klas	Stare klasy z defektami [%]	Nowe klasy z defektami [%]	Liczba przebadanych projektów
60019	9722	16,04	2,15	18



Liczba defektów naprawionych w poprzednim wydaniu projektu (NDPV)

Średnia	Odchylenie standardowe	Korelacja Pearsona z liczbą defektów	Korelacja Spearmana z liczbą defektów	Liczba przebadanych projektów
0,26	1,02	0,36	0,24	17





Korelacje pomiędzy różnymi metrykami oraz defektami

	NR	NDC	NML	NDPV	Defekty
NR	1.00	.71	.61	.53	.31
NDC	.71	1.00	.49	.43	.27
NML	.61	.49	1.00	.41	.27
NDPV	.53	.43	.41	1.00	.36
WMC	.33	.19	.15	.25	.14
DIT	-.08	-.11	-.05	-.05	-.08
NOC	.01	.01	.00	.01	.05
CBO	.23	.20	.23	.25	.10
RFC	.39	.28	.30	.35	.17
LCOM	.15	.04	.03	.12	.09
CA	.21	.16	.14	.17	.09
CE	.14	.13	.23	.23	.07
NPM	.27	.16	.09	.21	.16
LCOM3	-.09	-.07	-.05	-.07	-.09
LOC	.33	.24	.32	.30	.21
DAM	.16	.14	.05	.06	.11
MOA	.22	.13	.03	.13	.16
MFA	-.08	-.09	-.06	-.05	-.07
CAM	-.22	-.23	-.12	-.16	-.17
IC	-.05	-.08	-.04	-.02	-.06
CBM	.02	-.02	-.01	.02	-.03
AMC	.08	.12	.16	.10	.14
Max(CC)	.22	.19	.20	.19	.14
Min(CC)	.13	.14	.14	.13	.13



Agenda

- Wprowadzenie do modeli predykcji defektów
- Definicje metryk
- Metryki procesu w modelach predykcji defektów
- Statystyki opisowe
- Podsumowanie



Podsumowanie

- Metryki procesu są obiektem badań wielu prac dotyczących predykcji defektów.
- Większość badaczy rekomenduje stosowanie metryk procesu.
- Większość prac bazuje na małych zbiorach danych co ogranicza ich walidację zewnętrzną.
- Przytoczony eksperyment potwierdza użyteczność metryk procesu ale sygnalizuje potencjalne problemy związane z ich stosowaniem.



Politechnika Wroclawska

Dziękuję za uwagę