

Indeks	Punkty	Uwagi - sprawozdanie	Punkty (prezenta	Uwagi - prezentacja
258554	2	<ul style="list-style-type: none"> - dobra eksploracja danych - odnośnie usuwania zmiennych z wysokim p-value. Nie należy tego robić w „jednym ruchu”. Jeśli dwie zmienne są ze sobą mocno skorelowane (liniowo zależne), to dla obu takich zmiennych p-wartości mogą być duże, mimo że mają one istotny wkład w dopasowanie modelu - model nie ma sensu, bo zmienne jakościowe zostały uznane za ilościowe. Przecież Virginia nie jest dwa razy bardziej niż Luisiana. Dlatego AUC wyszło niskie. Accuracy jest dobre, bo zbiór jest niezbalansowany, ale porównując do baseline model (czyli wszystkich wypuszczamy) radzimy sobie gorzej 	4	<ul style="list-style-type: none"> - To prezentacja bardziej odnośnie tego co na wcześniejszych prezentacjach, a nie na temat - W USA nie chcą wypuszczać kogo popadnie, bo mają inne podejście do prawa. Prawo jest świetne! - Główna idea: nie pełny model tylko poprawka na to co zdecydują sędziowie, jest ok. Zwłaszcza w kontekście podejrzanego obciążenia próby
258486	4	<ul style="list-style-type: none"> - eksploracja danych + uwarunkowania prawne bardzo dobre. Sporo odważnych, ale uzasadnionych decyzji przy wyrzucaniu i kategoryzacji zmiennych - dlaczego lasso dopasowujecie bez interceptu? - dlaczego wybraliście właśnie ten model? AUC ma najniższe. Co w kształcie ROC sprawiło, że uznaliście ją za najlepszą? 	5	<ul style="list-style-type: none"> - Cel zadania opisany. - Co w ogóle mamy w danych? Opis problemów jest ok. - Dodanie nowych zmiennych - dobrze wyjaśnione. - zależność od stanu bardzo duża - zestawienie z danymi zewnętrznymi super. - Zestawienia z założeniami systemu prawnego (chcemy zwalniać ludzi, nie możemy patrzeć na płeć i rasę) - Fajne uwagi odnośnie ograniczeń modelu.
258536	2	<ul style="list-style-type: none"> - nie ma sensu liczyć korelacji pomiędzy zmiennymi czynnikowymi! - odnośnie usuwania zmiennych z wysokim p-value. Nie należy tego robić w „jednym ruchu”. Jeśli dwie zmienne są ze sobą mocno skorelowane (liniowo zależne), to dla obu takich zmiennych p-wartości mogą być duże, mimo że mają one istotny wkład w dopasowanie modelu - model nie ma sensu, bo zmienne jakościowe zostały uznane za ilościowe. Przecież Virginia nie jest dwa razy bardziej niż Luisiana. Dlatego AUC wyszło niskie. Accuracy jest dobre, bo zbiór jest niezbalansowany, ale porównując do baseline model (czyli wszystkich wypuszczamy) radzimy sobie gorzej 	4	<ul style="list-style-type: none"> - wykresy są nieczytelne, bo zmienne czynnikowe traktowane są jak ni
258601	4	<ul style="list-style-type: none"> - eksploracja danych + uwarunkowania prawne bardzo dobre. Sporo odważnych, ale uzasadnionych decyzji przy wyrzucaniu i kategoryzacji zmiennych - dlaczego lasso dopasowujecie bez interceptu? - dlaczego wybraliście właśnie ten model? AUC ma najniższe. Co w kształcie ROC sprawiło, że uznaliście ją za najlepszą? 	5	<ul style="list-style-type: none"> - Cel zadania opisany. - Co w ogóle mamy w danych? Opis problemów jest ok. - Dodanie nowych zmiennych - dobrze wyjaśnione. - zależność od stanu bardzo duża - zestawienie z danymi zewnętrznymi super. - Zestawienia z założeniami systemu prawnego (chcemy zwalniać ludzi, nie możemy patrzeć na płeć i rasę) - Fajne uwagi odnośnie ograniczeń modelu.

258544	3.5	<ul style="list-style-type: none"> - „być może za chwilę do czegoś dojdziemy." Raport ma być podsumowaniem, albo dojdziemy, albo nie - podział na długość odsiadki bardzo fajny - jak wybierane były zmienne do modeli? - w funkcji step.AIC można wyłączyć wypisywanie kroków pośrednich przez R - Podsumowanie nie jest dobre. Coś musi zostać zaproponowane. Nieprecyzyjne słownictwo typu „skombinować je jakoś ze sobą" jest niedopuszczalne. Po przeczytaniu podsumowania powinno być jasne co zostało zrobione i jak dobre są wyniki. Ani jednego, ani drugiego tutaj nie ma 		<ul style="list-style-type: none"> - Uwaga: przedstawienie dwóch modeli, ale który wybrać? Po co przedstawiać oba? - Model prosty wcale nie jest prosty - mamy zmienne jakościowe zamiast ilościowych, więc więcej parametrów - Droga dojścia do wyniku zamiast wyniku.
258609				
258512				
	4	<ul style="list-style-type: none"> - odnośnie usuwania zmiennych z wysokim p-value. Nie należy tego robić w „jednym ruchu". Jeśli dwie zmienne są ze sobą mocno skorelowane (liniowo zależne), to dla obu takich zmiennych p-wartości mogą być duże, mimo że mają one istotny wkład w dopasowanie modelu - AIC ok, ale brakuje próby zbudowania modelu LASSO - spadek sensitivity wynikać może z tego, że podział na zbiór tren, val, i test nie zachowuje proporcji naruszenie zwolnień w całej próbie - różnica to 10% - wybór punktu odcięcia powinien być zrobiony na zbiorze treningowym!!! Wtedy możemy sprawdzić czy nasze rozumowanie było poprawne 		<ul style="list-style-type: none"> ok 5 , bez uwag
258478	4.5	<ul style="list-style-type: none"> - plus za spis treści po lewej stronie dokumentu - w markdownie w R po rysunkach i przed akapitami trzeba dać dodatkową linię przerwy - plus za próbę zrozumienia różnic między stanami - nie „model całkiem prosty" tylko model pełny - wybór prostszego modelu jest ok. Brakuje jednak regresji LASSO - polecam przetwarzanie ramek danych w pakiecie dplyr, jest znacznie bardziej czytelny - nie ma potrzeby tłumaczyć w raporcie jak działają współczynniki - próba dodania zmiennych bardzo dobra 		<ul style="list-style-type: none"> - Kilka możliwości wykorzystania modelu w zależności od tego, co chce się uzyskać. - Opis modelu w porządku. - Wyłumaczenie wpływu zmiennych w modelu przez barploty - ok. - Wyłumaczenie prostoty modelu. Brakuje odniesienia do modelu sędziowskiego. 4 - Możliwości rozszerzenie i problemy z użyciem.
<p>Uwagi ogólna: Nie zawsze trzeba wypisywać całej informacji o dopasowaniu modelu. Do selektywnego udostępniania informacji polecam pakiet broom</p>				