

# PD LOOKUP

*Agata Buksińska*

*Aleksandra Chmarzyńska*

*Karolina Mucha*

# SPIIS TREŚCI

- 01 - Cel projektu
- 02 - Ryzyko w działalności banku
- 03 - Definicja *default* i *default rate*
- 04 - Model *Probability of default*
- 05 - Etapy budowy i weryfikacji modelu PD
- 06 - Dane i prawa autorskie



# CEL PROJEKTU

Budowa modelu scoringowego  
przewidującego zdarzenie wejścia  
klienta w default



# RYZYKO W DZIAŁALNOŚCI BANKU

W literaturze przedmiotu nie istnieje jednolita klasyfikacja ryzyka bankowego.

W dużej mierze zależy ona od działalności i profilu danego banku (można spotkać kilkadziesiąt podziałów opartych na różnych kryteriach).

Jednym z podziałów jest podział na ryzyko pierwotne (rozpatrywane z punktu widzenia przyczyn) oraz wtórne (z punktu widzenia skutków) wg A. Kuglera.

## RYZYKO PIERWOTNE:

ryzyko kredytowe,  
ryzyko stopy procentowej,  
ryzyko walutowe.

## RYZYKO WTÓRNE:

ryzyko utraty płynności,  
ryzyko nieosiągnięcia wyniku,  
ryzyko standingu.



# PODZIAŁ RYZYKA (KOMITET BAZYLEJSKI)

- ryzyko **kredytowe**;
- ryzyko **rynkowe**, w tym:
  - ryzyko walutowe,
  - ryzyko cen towarów,
  - ryzyko cen kapitałowych papierów wartościowych,
  - ryzyko szczególne cen instrumentów dłużnych,
  - ryzyko ogólne stóp procentowych;
- ryzyko **operacyjne** (w tym ryzyko prawne).

# RYZYKO KREDYTOWE

Ryzyko kredytowe oznacza niebezpieczeństwo, iż kredytobiorca nie wypełni zobowiązań i warunków zawartych z bankiem umowy, narażając kredytodawcę na powstanie straty finansowej.

Ryzyko kredytowe jest jednym z najważniejszych w banku (za wyjątkiem banków inwestycyjnych), ponieważ odnosi się do sytuacji zagrażających egzystencji banku (utrata zdolności do terminowego wywiązywania się ze zobowiązań oraz powstanie strat finansowych pochłaniających fundusze własne, co grozi upadłością).

**Ryzyko aktywne (czynne)**

**Ryzyko pasywne (bierne)**



NON-DEFAULT



DEFAULT

# DEFAULT

**Default** to stan, w którym **klient banku wykazuje wyraźne oznaki utraty zdolności do regulowania swoich kredytowych zobowiązań.**

Istnieje wysokie ryzyko, że nie będzie w stanie wywiązać się ze swoich długów wobec instytucji finansowej.

Definicja defaultu to kryteria i warunki, na podstawie których określa się, że klient znajduje się w stanie niewywiązania się z zobowiązań kredytowych.

# NAJISTOTNIEJSZE PARAMETRY RYZYKA

## **Prawdopodobieństwo Niewykonania Zobowiązań (PD):**

Ten wskaźnik jest fundamentem analizy ryzyka kredytowego, reprezentując szacunkową możliwość, że kredytobiorca nie będzie w stanie spłacić długu w przewidywanym okresie. PD jest wyrażane jako procent i bazuje na historycznych danych, obecnej kondycji finansowej kredytobiorcy oraz czynnikach zewnętrznych wpływających na zdolność do spłaty.

To wskaźnik, który dynamicznie reaguje na zmiany w środowisku gospodarczym i osobistej sytuacji kredytobiorcy, stanowiąc kluczowy element w zarządzaniu portfelem kredytowym banku.

## **Eksponowanie Zagrożone (EAD):**

EAD określa całkowitą kwotę, która mogłaby być zagrożona w przypadku niewykonania zobowiązań. Jest to suma aktualnie wykorzystanego kredytu oraz potencjalnie dodatkowej kwoty, która może zostać wypłacona przed domyślnym terminem.

## **Strata w Przypadku Niewykonania Zobowiązań (LGD):**

LGD to estymacja procentowej straconej wartości, gdy dłużnik nie spłaci długu. Jest to liczona po uwzględnieniu ewentualnych zabezpieczeń, które bank może zrealizować.

**Wzór na Oczekiwanyą Stratę (ECL):**  $ECL = PD \times EAD \times LGD$

# DEFAULT RATE

## Jest podstawowym miernikiem PD

Oznacza stosunek liczby przypadków niewykonania zobowiązania, do których doszło w okresie obejmującym dany okres (często 12 msc) poprzedzający datę T, do liczby dłużników zaklasyfikowanych do tej klasy lub puli na dany okres przed upływem tej daty.

$$\text{DR}_{\text{jednoroczny}} =$$

$$\frac{\text{Liczba dłużników spośród wszystkich dłużników uwzględnionych w mianowniku, którzy mieli przynajmniej jedno zdarzenie niewykonania zobowiązania w jednorocznym okresie obserwacji}}{\text{Liczba dłużników wykonujących swoje zobowiązanie, przy czym zobowiązanie jest obserwowane na początku jednorocznego okresu obserwacji}}$$

# Model PD Prawdopodobieństwa Niewykonania



**Model PD to narzędzie finansowe wykorzystywane do oszacowania ryzyka, że kredytobiorca nie spłaci swojego zadłużenia.** Jest to wyrażone w procentach prawdopodobieństwo, że dłużnik znajdzie się w sytuacji, w której nie będzie w stanie wywiązać się ze swoich finansowych zobowiązań w określonym przedziale czasowym. Model ten bazuje na analizie historycznych danych dotyczących niewykonania zobowiązań przez kredytobiorców, co pozwala na wypracowanie miernika ryzyka nazywanego wskaźnikiem PD.

## Kluczowe cechy modelu PD:

**Oparte na danych:** Model PD wykorzystuje historyczne przypadki defaultu (niewykonania zobowiązań) do szacowania przyszłego ryzyka niewykonania.

**Dynamiczny:** Model jest aktualizowany o nowe dane, aby odzwierciedlać zmieniające się warunki rynkowe i indywidualne sytuacje kredytobiorców.

**Wskaźnik Default Rate:** To obserwowana częstotliwość defaultu, która jest podstawą do wyliczenia PD.



# DO CZEGO SŁUŻY MODEL PD?

- 1) Wyznaczanie odpisów z tytułu utraty wartości (rezerwy) – oczekiwana strata kredytowa
- 2) Wyznaczanie wartości wymogów kapitałowych (IRB) – nieoczekiwana strata
- 3) Prowadzenie/wspieranie działalności kredytowej – ocena ryzyka przy przyznawaniu kredytów

# METODY MODELOWANIA PD

**Macierze migracji** to narzędzia używane w analizie ryzyka kredytowego do śledzenia zmian jakości portfela kredytowego w czasie. Pokazują, jakie kategorie ryzyka kredytowego (np. ratingi) zmieniają się dla danej grupy klientów w kolejnych okresach. Założenie własności Markova, w kontekście macierzy migracji często jest nieprawdziwe - przyszłe zmiany w kategorii ryzyka zależą tylko od obecnej kategorii ryzyka, a nie od całej historii zmian.

**Regresja logistyczna** jest to statystyczna metoda używana do modelowania zależności pomiędzy zmienną zależną, która jest binarna (default klienta), a jednym lub wieloma predyktorami. Jest często stosowana w branży bankowej do budowy modeli scoringowych i prognozowania prawdopodobieństwa defaultu (PD). Regresja logistyczna, choć powszechnie stosowana w bankowości, może mieć ograniczone możliwości predykcyjne w porównaniu z bardziej zaawansowanymi metodami uczenia maszynowego (ML).

**Metody typu vintage** są to metody analizy portfela kredytowego, które biorą pod uwagę czas, w jakim dany kredyt został udzielony. Działają na danych podzielonych według okresu otwarcia kredytu, co umożliwia analizę wyników dla różnych "roczników" kredytów. Metody typu vintage są łatwe technicznie, ale mogą być trudne w segmentacji, ponieważ wymagają dużego zbioru danych, a dokładne wyniki są często uzyskiwane dopiero po długim okresie obserwacji.

# MODEL REGRESJI LOGISTYCZNEJ

**Regresja logistyczna** umożliwia prognozowanie prawdopodobieństwa danego zdarzenia na podstawie zbioru danych, co pozwala podejmować decyzje i podejście do zarządzania ryzykiem.

$$P(Y = 1) = \frac{e^{\alpha_0 + \sum_{i=1}^k \alpha_i x_i}}{1 + e^{\alpha_0 + \sum_{i=1}^k \alpha_i x_i}}$$

gdzie:

$P(Y = 1)$  to prawdopodobieństwo **defaultu**,  
 $\alpha_0, \alpha_1, \dots, \alpha_n$  to współczynniki regresji,  
 $x_1, x_2, \dots, x_n$  to wartości **predyktorów**.

# ZAŁOŻENIA REGRESJI LOGISTYCZNEJ

- W analizie regresji logistycznej, interesuje nas prognozowanie zdarzenia, które może przyjąć tylko jedną z dwóch kategorii, na przykład sukces lub niepowodzenie, tak jak w przypadku defaultu klienta.
- W trakcie budowy modelu regresji logistycznej ważne jest, aby uwzględnić wszystkie istotne zmienne, które wpływają na zmienną zależną, jednocześnie eliminując te, które nie mają istotnego wpływu na prognozowane zdarzenie.
- Proces transformacji logitowej w regresji logistycznej zachodzi w sposób liniowy w odniesieniu do zmiennych niezależnych, co oznacza, że wpływ każdej zmiennej jest proporcjonalny do jej wartości.
- Regresja logistyczna nie uwzględnia automatycznie interakcji między zmiennymi niezależnymi, co oznacza, że efekty współdziałania między nimi nie są bezpośrednio uwzględniane w modelu.
- W analizie regresji logistycznej ważne jest, aby unikać współliniowości między zmiennymi niezależnymi, czyli sytuacji, w której dwie lub więcej zmiennych są silnie skorelowane.
- Model regresji logistycznej może być podatny na wpływ punktów odstających, czyli obserwacji o nietypowych wartościach, co może wpływać na stabilność i dokładność prognoz.
- Aby uzyskać wiarygodne i stabilne wyniki regresji logistycznej, próba powinna być odpowiednio liczna, z zaleceniem, że liczba obserwacji w każdej klasie powinna być co najmniej 10 razy większa niż liczba zmiennych wyjaśniających w modelu, czyli np. przy 10 zmiennych powinno być minimum 100 obserwacji w klasie default i 100 w klasie non-default.



Do uzyskania kompleksowej oceny wydajności modelu używa się miar predykcji:

- **Dokładność** (Accuracy), jako procent poprawnych predykcji w stosunku do wszystkich przypadków.
- **Czułość** (Sensitivity, Recall, True Positive Rate), jako procent poprawnych predykcji dla pozytywnych przypadków rzeczywistych.
- **Specyficzność** (Specificity, True Negative Rate), jako procent poprawnych predykcji dla negatywnych przypadków rzeczywistych.
- **Precyzja** (Precision, Positive Predictive Value), jako procent poprawnych predykcji pozytywnych w stosunku do wszystkich pozytywnych predykcji.
- **F1 Score**, czyli ważona średnia harmoniczna precyzji i czułości.
- **Krzywa ROC**, czyli graficzna reprezentacja wydajności modelu w różnych punktach odcięcia.
- **Obszar pod krzywą ROC** (AUC-ROC) ocenia ogólną wydajność modelu niezależnie od punktu odcięcia.
- **Miara dokładności wieloklasowej**, jako procent poprawnych predykcji dla wszystkich klas.

Do uzyskania kompleksowej oceny wydajności modelu używa się miar predykcji:

- **Średni Błąd Kwadratowy** (Mean Squared Error - MSE), czyli średnia kwadratów różnic między przewidywanymi a rzeczywistymi wartościami.
- **Średni Błąd Bezwzględny** (Mean Absolute Error - MAE), czyli średnia z bezwzględnych różnic między przewidywanymi a rzeczywistymi wartościami.
- **R<sup>2</sup>** (Coefficient of Determination) mierzy odsetek wariancji w danych zmiennych zależnych, który jest wyjaśniany przez model.
- **Błąd Procentowy Średni** (Mean Percentage Error - MPE), czyli średnia z procentowych różnic między przewidywanymi a rzeczywistymi wartościami.
- **Błąd Procentowy Bezwartościowy** (Mean Absolute Percentage Error - MAPE), czyli średnia z bezwzględnych procentowych różnic między przewidywanymi a rzeczywistymi wartościami.
- **Logarithmic Loss** mierzy jakość klasyfikacji prawdopodobieństw. Im niższa wartość, tym lepiej.

gdzie:

N to liczba obserwacji,

$y_i$  to rzeczywista klasa (0 lub 1),

$p_i$  to przewidziane prawdopodobieństwo

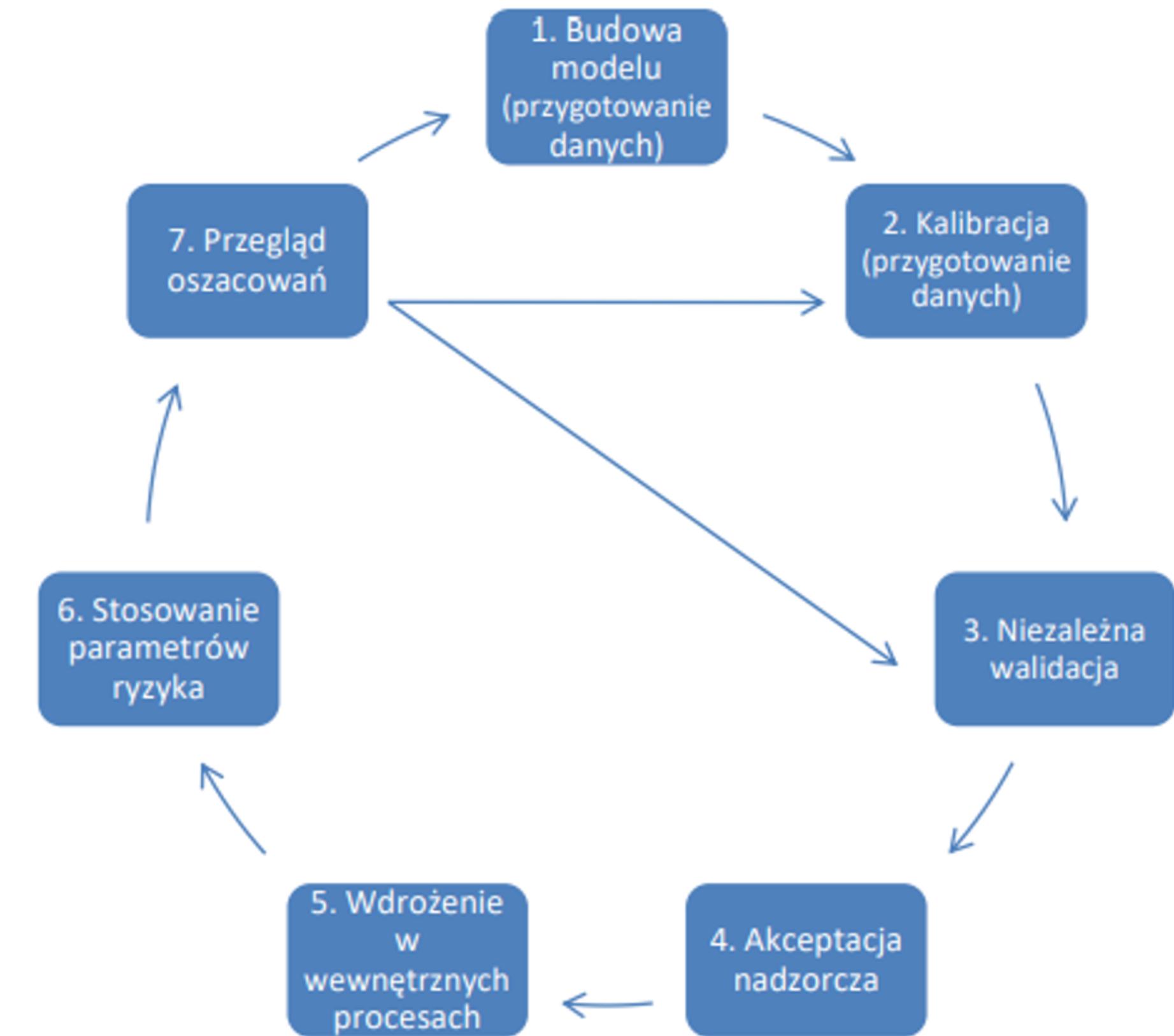
$$LogLoss = -\frac{1}{N} \sum_{i=1}^N [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)],$$

- **Brier Score** mierzy dokładność prognoz prawdopodobieństw. Im niższa wartość, tym lepiej:

$$Brier Score = \frac{1}{N} \sum_{i=1}^N (y_i - p_i)^2$$

# CYKL ŻYCIA MODELU

Cykl życia modelu jest iteracyjny, a fazy mogą się powtarzać w zależności od potrzeb biznesowych i zmieniających się warunków. Monitoring i walidacja odgrywają kluczową rolę w zapewnieniu, że model jest skuteczny i zgodny z oczekiwaniami, nawet po jego wdrożeniu.



# DANE DO MODELU

Zestaw danych będzie się składał z kilkudziesięciu tysięcy ekspozycji kredytowych oraz wybranych danych potrzebnych do budowy modelu.

Zestaw danych zostanie podzielony jest na dwie próbki;

70% – próba ucząca,  
30% – próba testowa



# BIBLIOGRAFIA

ZALESKA, MAŁGORZATA, ED. WSPÓŁCZESNA BANKOWOŚĆ. CENTRUM DORADZTWA I INFORMACJI " DIFIN", 2007, S. 278-281

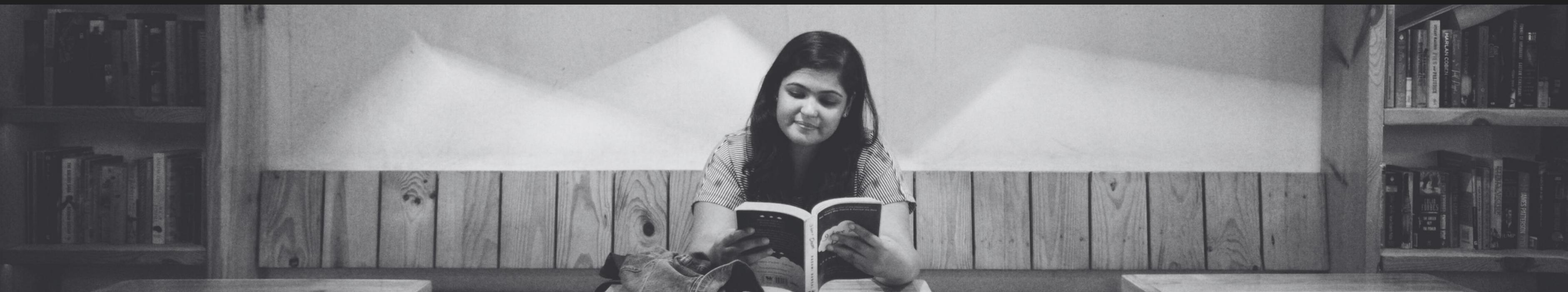
UCHWAŁA KNB NR 1/2007 Z DNIA 13.03.2007 R. (DZIENNIK URZĘDOWY NBP NR 2/2007 POZ. 3). RYZYKO W DZIAŁALNOŚCI BANKOWEJ: ZABEZPIECZENIA SYSTEMOWE, SZKOŁA GŁÓWNA HANDLOWA, WARSZAWA 1992, S. 14.

BUCZYŃSKI M., LEWANDOWSKI B., PRAKTYCZNE ASPEKTY IMPLEMENTACJI WYTYCZNYCH EBA, W TYM DOTYCZĄCYCH BUDOWY MODELI PD/LGD ORAZ STOSOWANIA DEFINICJI DEFAULT, KNF, 2018 [HTTPS://WWW.KNF.GOV.PL/KNF/PL/KOMPONENTY/IMG/PREZENTACJA%20CEDUR%2014-12-2018\\_64235.PDF](https://www.knf.gov.pl/knf/pl/komponenty/img/prezentacja%20cedur%2014-12-2018_64235.pdf)

[HTTPS://WWW STOCKWATCH PL/WIADOMOSCI/ZROZUMIEC-ODPISY-KREDYTOWE-CO-ZMIENIA-MSSF-9,ANALIZY,217194](https://www.stockwatch.pl/wiadomosci/zrozumieć-odpisy-kredytowe-co-zmienia-mssf-9,analizy,217194)

[HTTPS://BAZEKON.UEK.KRAKOW.PL/ZESZYTY/12843](https://bazekon.uek.krakow.pl/zeszyty/12843)

[HTTPS://WWW.KNF.GOV.PL/KNF/PL/KOMPONENTY/IMG/BANKI\\_MAPA\\_RZYK\\_25314.PDF](https://www.knf.gov.pl/knf/pl/komponenty/img/banki_mapa_rzyk_25314.pdf)



DZIĘKUJEMY

*za uwagę*