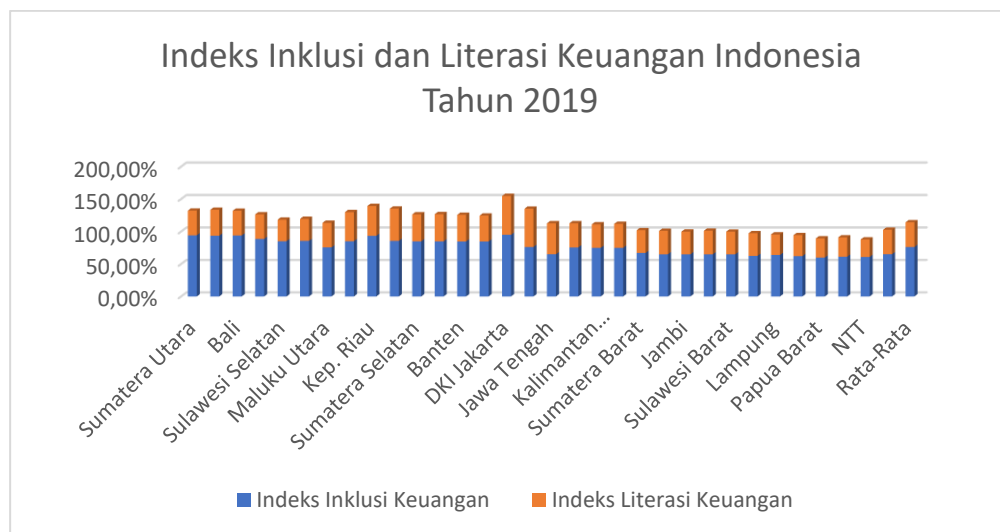


# BAB 1 PENDAHULUAN

## 1.1 Latar Belakang

Menurut laporan (KPMG, 2020) berjudul *The Pulse of Fintech 2018* menyebutkan bahwa Asia Tenggara menjadi target pertumbuhan *fintech* dan ekspansi global. Banyak perusahaan *fintech* terutama perusahaan *fintech* yang berasal dari China melihat Asia Tenggara sebagai pasar yang besar. Beberapa perusahaan besar China melakukan investasi secara global di banyak negara. Mayoritas bank di China juga mulai fokus pada digital dan telah menyebabkan pertumbuhan pesat bisnis berorientasi *fintech*. Mereka berinvestasi di banyak bidang termasuk data besar, *blockchain*, dan kecerdasan buatan. Fenomena ini tidak mengherankan karena China memiliki pertumbuhan layanan keuangan digital yang fenomenal (Zhou et al., 2018) dengan jumlah *P2P Lending* terbesar di dunia (Stern et al., 2017)



Gambar 1.1 Indeks Inklusi dan Literasi Keuangan Indonesia Tahun 2019  
Berdasarkan Pembagian Provinsi (OJK, 2021)

Indonesia merupakan bagian dari Asia Tenggara yang memiliki literasi dan inklusi keuangan yang rendah. Survei Nasional Literasi dan Inklusi Keuangan dari Otoritas Jasa Keuangan (2019) sesuai pada Gambar 1.1 menunjukkan bahwa tingkat literasi keuangan adalah 38,03 persen dan tingkat inklusi keuangan adalah 76,16 persen. Tingkat inklusi keuangan yang lebih tinggi menunjukkan bahwa

masyarakat menggunakan jasa keuangan tanpa didasari oleh pengetahuan keuangan yang memadai. Dalam survei tersebut, ditemukan bahwa literasi dan inklusi masyarakat tentang perbankan masih mendominasi dibandingkan dengan layanan keuangan lainnya.

Indonesia sebagai negara yang memiliki penduduk terbesar di Asia memiliki potensi besar bermunculan berbagai *platform Financial Technology (Fintech)* antara lain *peer-to-peer (P2P) lending*, *Aggregator*, *Project Financing*, *Blockchain-based*, *Funding Agent*, *Insurtech*, dan *Financing Agent*.

Kajian *Institute for Development of Economics and Finance (INDEF)* bersama *Association of Fintech Indonesia (AFTECH)* menyimpulkan bahwa *P2P lending* memberikan kontribusi positif, yakni menyumbang Rp 25,97 triliun terhadap PDB di Indonesia, dimana sektor yang memberikan kontribusi terbesar adalah jasa keuangan sebesar menjadi Rp 4,7 triliun, menyalurkan pinjaman sebesar Rp 7,64 triliun kepada 1,47 juta orang dan menciptakan lapangan kerja bagi lebih dari 350.000 orang di berbagai sektor.

Di Indonesia seiring bermunculnya *P2P lending* legal, muncul juga *P2P lending* ilegal yang dimana masih belum mendapat izin dari pihak berwenang, yaitu setidaknya ada 400 perusahaan *fintech* ilegal yang beroperasi di Indonesia. Keberadaan *fintech* ilegal menimbulkan banyak masalah, terutama bagi peminjam yang terlambat membayar. Selama tahun 2018, terdapat lebih dari 1.300 pengaduan ke Lembaga Bantuan Hukum Jakarta (LBH Jakarta) terkait kasus pinjaman *online* (Hidajat, 2020). Jumlah ini diduga masih kecil karena belum semua korban melapor. Keberadaan *fintech* ilegal merugikan karena berpotensi menjadi media pencucian uang, pendanaan terorisme, penggelapan pajak, penyalahgunaan data pelanggan, mengurangi kepercayaan terhadap *peer-to-peer lending* legal, dll.

Industri P2P Lending di Indonesia menghadapi tantangan yang tak kalah menarik dibandingkan dengan metode tradisional dalam membentuk kredit. P2P Lending di Indonesia masih mengalami kesulitan dalam mendapatkan data kredit atau informasi keuangan konsumen, yang berbeda dengan lembaga keuangan lainnya. Hal ini mengakibatkan keterbatasan data dalam menilai kelayakan kredit (Guo et

al., 2016). Karena keterbatasan data yang tersedia, perusahaan P2P Lending dihadapkan pada tantangan untuk mencari alternatif data atau informasi tambahan yang dapat digunakan sebagai penilaian kredit. Oleh karena itu, banyak perusahaan P2P Lending menggunakan data dari media sosial sebagai salah satu sumber informasi (O’Leary, Steve. Sheehan, Kim. Sterling, 2013), Dengan memanfaatkan informasi dinamis yang ada di media sosial, risiko dan kerugian dalam memberikan kredit dapat dikendalikan. Namun, data dari media sosial tidak secara jelas menggambarkan konten terkait masalah kredit, seperti dalam teks pendek yang seringkali tidak mencakup data statistik yang memadai untuk mendukung berbagai permasalahan (Hua et al., 2015). Pandangan tersebut juga didukung oleh Guo et al., (2016) yang menyatakan bahwa data dari media sosial cenderung memberikan informasi yang tidak relevan terkait dengan kredit, sehingga menjadi tantangan besar dalam menentukan kelayakan kredit.

Adapun beberapa penelitian sebelum *machine learning* menjadi populer dalam klasifikasi *credit scoring*, yaitu banyak penelitian dilakukan dengan menggunakan metode statistik tradisional dan analisis data konvensional yaitu dengan menggunakan analisis diskriminan (Eisenbeis, 1978) pada penelitian tersebut menggunakan analisis diskriminan untuk memisahkan data pelanggan menjadi kelompok kredit yang berbeda berdasarkan variabel-variabel seperti pendapat, riwayat kredit, dan rasio utang. Kemudian adapun penelitian lainnya yaitu seperti menggunakan regresi logistik (Wiginton, 1980) yang dimana penelitian tersebut fokus pada pengembangan model regresi logistik yang memprediksi probabilitas keterlambatan pembayaran atau kegagalan pembayaran berdasarkan berbagai variabel yang relevan.

Algoritma *NBC* (*Naïve Bayesian Classification*) merupakan algoritma klasifikasi yang sangat efektif (mendapatkan hasil yang tepat) dan efisien (proses penalaran dilakukan memanfaatkan input yang ada dengan cara yang relatif cepat). Algoritma *NBC* bertujuan untuk melakukan klasifikasi data pada kelas tertentu. Hasil kerja pengklasifikasian diukur dengan nilai *predictive accuracy* (H. Zhang & Su, 2004). Model statistik merupakan salah satu model yang terpercaya sangat handal sebagai pendukung pengambilan keputusan. Konsep probabilitas merupakan salah satu bentuk model statistik. Salah satu

metode yang menggunakan konsep probabilitas adalah *Naive Bayesian Classification (NBC)*. Pada metode ini, semua atribut akan memberikan kontribusinya dalam pengambilan keputusan, dengan bobot atribut yang sama penting dan setiap atribut saling bebas satu sama lain (Kusumadewi, 2009).

Menurut penelitian (Vedala & Kumar, 2012), penelitian dilakukan dengan mengambil data dari calon peminjam tertentu dengan berdasarkan profilnya dan informasi yang diambil dari jejaring sosialnya seperti profil dan aktivitas teman peminjam (seperti aktivitas peminjaman, peminjaman, dan penawaran), dan menggunakan algoritma klasifikasi *Naïve Bayes Classification* untuk mengklasifikasikan peminjam yang baik dan buruk, di mana atribut input terdiri dari informasi kredit inti dan jaringan sosial.

Oleh karena itu peneliti melakukan penelitian ini dengan menggunakan metode *Naïve Bayes Classification* yang dimana metode ini sudah sering dipakai dibandingkan dengan metode klasifikasi lainnya, dan juga metode ini memiliki tingkat efektivitas yang tinggi dan dapat bekerja pada tipe variabel input diskrit atau kategoris yang dapat dilakukan pada penelitian ini.

Kemudian dalam membuat model prediksi *credit scoring* pada penelitian ini akan menggunakan *dataset* dari lendingclub.com yang dimana sudah mengandung record data-data kredensial seperti pendapatan tahunan, status pekerjaan, status tempat tinggal, data historis dari *customer* dari platform *peer-to-peer lending* lendingclub.

## **1.2 Perumusan Masalah**

Berdasarkan uraian masalah yang telah dijelaskan pada latar belakang diatas, maka terlihat masalah yang ditemukan pada penelitian ini dan penulis dapat merumuskan masalah sebagai berikut :

1. Model algoritma klasifikasi yang tepat akan dapat memprediksi tingkat *credit score* pada *platform peer-to-peer lending* yang sedang dikembangkan.
2. Tingkat akurasi yang tinggi oleh model algoritma akan diimplementasikan akan dapat memberikan keputusan yang tepat bagi platform *peer-to-peer*

*lending* yang sedang dikembangkan dalam memberi pinjaman bagi calon peminjam.

3. Dalam menentukan *credit scoring* pada platform *peer-to-peer lending*, aplikasi membutuhkan sebuah model yang dapat terintegrasi dengan aplikasi yang dikembangkan dalam mengetahui *credit score* pada suatu pinjaman.

### 1.3 Tujuan Penelitian

Berdasarkan latar belakang dan rumusan masalah yang sudah dijelaskan diatas, tujuan yang ingin dicapai dari penelitian ini adalah :

1. Mengimplementasikan Model *Machine Learning* menggunakan algoritma *Naïve Bayesian Classification* untuk memprediksi klasifikasi kredit skor dari peminjam.
2. Melakukan *hyperparameter tuning* terhadap tingkat akurasi model prediksi yang diimplementasikan.
3. Melakukan *deploy* API dari model algoritma *Naïve Bayes Classifier* pada *Server Google Cloud Platform*.

### 1.4 Batasan Penelitian

Adapun batasan dalam penelitian ini, adalah sebagai berikut :

1. *Dataset* yang dipakai pada penelitian ini disadur dari [www.lendingclub.com](http://www.lendingclub.com) yang merupakan salah satu *platform peer-to-peer lending* yang berasal dari Inggris.
2. Atribut yang digunakan pada penelitian ini akan dipilih berdasarkan hasil korelasi linear semua atribut terhadap atribut *grade* kemudian akan diseleksi kembali dengan kesesuaian atribut yang ada pada platform *peer-to-peer lending* yang sedang dikembangkan.
3. Pembagian *Data Training* dan *Data Testing* akan terbagi menjadi dua tahap simulasi yang akan digunakan pada pengujian atau evaluasi yang dimana berfungsi untuk mencapai nilai *accuracy score* yang maksimal dan setiap tahap memiliki tiga simulasi *data splitting* yaitu tahap pertama yaitu *data splitting* untuk tahap sebelum melakukan *hyperparameter tuning* pada algoritma *Naïve Bayes* yang terdiri dari :

- a. Simulasi ke-1 pembagian *Data Training* 70% dan *Data Testing* 30%
- b. Simulasi ke-2 pembagian *Data Training* 80% dan *Data Testing* 20%
- c. Simulasi ke-3 pembagian *Data Training* 90% dan *Data Testing* 10%

Adapun tahap kedua yaitu melakukan *data splitting* untuk melakukan *hyperparameter tuning* pada algoritma *Naïve Bayes* yang terdiri dari :

- a. Simulasi ke-1 pembagian *Data Training* 70% dan *Data Testing* 30% kemudian dilakukan *data splitting* kembali menjadi variabel *Data Validation* 50% dari *Data Testing*.
- b. Simulasi ke-2 pembagian *Data Training* 80% dan *Data Testing* 20% kemudian dilakukan *data splitting* kembali menjadi variabel *Data Validation* 50% dari *Data Testing*.
- c. Simulasi ke-3 pembagian *Data Training* 90% dan *Data Testing* 10% kemudian dilakukan *data splitting* kembali menjadi variabel *Data Validation* 50% dari *Data Testing*.

## 1.5 Manfaat Penelitian

Hasil dari penelitian ini diharapkan memberikan manfaat, baik secara teoritis maupun praktis, diantaranya sebagai berikut :

1. Manfaat bagi industri, diharapkan dengan adanya penelitian ini dapat menjadi referensi bagi industri *peer-to-peer lending* lainnya dalam pengembangan platformnya.
2. Manfaat bagi masyarakat, diharapkan dengan adanya penelitian ini dapat mengurangi sistem tidak etis yang telah dilakukan oleh *peer-to-peer lending illegal* yang telah merugikan masyarakat.
3. Manfaat bagi peneliti, diharapkan dengan adanya penelitian ini peneliti dapat memahami bahwa teknologi keuangan (*fintech*) *lending* ataupun konvensional sangat membantu masyarakat dalam memenuhi kebutuhan primer ataupun sekundernya.

## 1.6 Sistematika Penelitian

Sistematika penulisan ini terbagi menjadi beberapa bab pokok pembahasan, secara umum dijabarkan sebagai berikut :

- a) BAB I – PENDAHULUAN, bab ini berisi tentang penjelasan mengenai latar belakang, rumusan masalah, tujuan penelitian, batasan penelitian, manfaat penelitian dan sistematika penelitian.
- b) BAB II – TINJAUAN PUSTAKA, bab ini berisi tentang penjelasan kajian kajian literatur pendukung untuk riset dan beberapa related work yang pernah dilakukan oleh peneliti sebelumnya.
- c) BAB III – METODOLOGI PENELITIAN, bab ini berisi tentang penjelasan mengenai konseptual model dan sistematika penelitian yang digunakan.
- d) BAB IV – ANALISIS DAN PERANCANGAN, bab ini berisi tentang perhitungan sebuah model analisis yang digunakan untuk penentuan faktor dari pengolahan data yang dilakukan dalam penelitian.
- e) BAB V – IMPLEMENTASI DAN HASIL EVALUASI, bab ini berisi tentang implementasi metode pengolahan data.
- f) BAB VI – KESIMPULAN DAN SARAN, bab ini berisi tentang kesimpulan hasil dari penelitian yang dilakukan dan saran yang dapat dipertimbangkan untuk penelitian selanjutnya.