

## 1. Pendahuluan

### Latar Belakang

Dalam penelitian ini, masalah yang dibahas adalah prediksi *return* saham menggunakan teknik pengelompokan K-Medoids dan model Long Short-Term Memory (LSTM) [1]. Tujuan utama adalah untuk memprediksi harga saham yang dikonversikan menjadi prediksi *return*, menggunakan pendekatan berbasis pengelompokan untuk mengidentifikasi saham dengan pergerakan harga yang sama. Pendekatan ini sangat penting karena memungkinkan investor untuk membuat keputusan investasi yang lebih baik dengan memilih saham yang memiliki potensi *return* tinggi dengan risiko yang lebih rendah. Pengelolaan aset keuangan yang efektif adalah kunci bagi investor yang berusaha untuk memaksimalkan hasil *return* sambil meminimalkan risiko. Optimalisasi portofolio telah menjadi pendekatan yang populer bagi investor, dengan model *mean-variance* (MV) yang diusulkan oleh Markowitz (1952) adalah contoh yang sukses. Meskipun telah terjadi kemajuan dalam optimasi portofolio, menggabungkan model *machine learning* (ML) dapat lebih meningkatkan kinerja [3]. Dengan menggunakan model prediksi untuk memprediksi *return* saham, investor dapat meningkatkan kinerja portofolio. *Return* merupakan salah satu faktor penentu utama bagi investor dalam berinvestasi dan merupakan imbalan atas keberanian mereka dalam menanggung risiko investasi. Dalam pembentukan portofolio, pertimbangan utama adalah *return* yang dihitung dari data historis. Kinerja portofolio dianggap baik jika *return* yang didasarkan pada data historis tidak jauh berbeda dengan *return* masa depan. Namun, dalam kenyataannya, data masa depan mungkin sangat berbeda dari data historis. Oleh karena itu, ketika membangun portofolio, perlu mempertimbangkan prediksi *return* di masa depan [2]. Dalam konteks ini, metode K-Medoids digunakan untuk mengelompokkan saham berdasarkan pola pergerakan harga yang sama, yang dapat membantu mengidentifikasi saham dengan kinerja yang mirip. Sementara itu, model LSTM digunakan untuk memprediksi harga yang dikonversikan menjadi *return* berdasarkan *clustering*, yang diharapkan dapat lebih akurat dalam memperkirakan *return*. *Return* saham merupakan faktor penentu utama dalam pengambilan keputusan investasi, dan dengan memanfaatkan prediksi *return* berbasis ML, investor dapat mengoptimalkan kinerja portofolio mereka. Masalah utama dalam penelitian ini adalah memprediksi harga saham yang diubah menjadi prediksi *return*, menggunakan teknik *clustering* untuk mengelompokkan saham berdasarkan pergerakan harga yang mirip, yang sangat berguna dalam mengatur data untuk *training* model LSTM. Masalah ini sangat relevan karena memungkinkan investor untuk membuat keputusan investasi yang lebih informasi dengan mengandalkan prediksi untuk kelompok saham tertentu.

Pada tahun 2023, Ashrafzadeh dkk melakukan penelitian menggunakan model LSTM untuk seleksi awal saham dan pengelompokan berbasis *k-means clustering*. Data yang digunakan terdiri dari 21 saham yang dipilih secara acak dari *New York Stock Exchange* (NYSE). Data ini mencakup periode harian selama sepuluh tahun, dari tahun 2012 hingga 2021. Saham-saham dikelompokkan berdasarkan fitur seperti rata-rata *return* dan standar deviasi *return*, dengan memanfaatkan nilai rata-rata dan variansi fitur input serta *return* aset selama periode *training* sebagai kriteria *clustering*. Model LSTM kemudian digunakan untuk memprediksi harga untuk masing-masing perwakilan *cluster*, yang kemudian dikonversi menjadi *return* dengan menghitung perubahan harga relatif terhadap harga sebelumnya. *Return* ini digunakan sebagai indikator kinerja saham dan diolah melalui model LSTM untuk mendapatkan prediksi *return* masa depan. Hasil penelitian menunjukkan bahwa model LSTM efektif dalam memprediksi dengan nilai RMSE untuk setiap *cluster*, yakni 0.0361, 0.0308, 0.0539, 0.0361, dan 0.0332, yang menunjukkan tingkat akurasi prediksi yang beragam dan bermanfaat untuk seleksi saham dalam portofolio investasi. Nilai rata-rata *return* LSTM adalah 0.003 dan standar deviasi *return* 0.0212, yang menegaskan bahwa model ini dapat digunakan untuk strategi investasi yang menguntungkan dalam jangka panjang [3].

Pada tahun 2022, Man Li dkk melakukan prediksi dengan menggunakan fitur rata-rata *return*, standar deviasi *return*, dan distribusi *return* harian untuk mengelompokkan saham dari *New York Stock Exchange* (NYSE) dalam periode 2012 hingga 2021. Model LSTM digunakan untuk memprediksi harga saham, yang kemudian dikonversi menjadi *return* dengan menghitung perubahan harga relatif terhadap harga sebelumnya. Hasil pengelompokan dan prediksi menunjukkan bahwa model LSTM memiliki nilai Root Mean Square Error (RMSE) yang bervariasi untuk setiap *cluster*, yaitu 0.0361, 0.0308, 0.0539, 0.0361, dan 0.0332, yang mengindikasikan tingkat akurasi prediksi yang berbeda-beda. Rata-rata *return* yang diprediksi oleh model LSTM adalah 0.003 dengan standar deviasi *return* sebesar 0.0212, menunjukkan bahwa model ini dapat digunakan untuk strategi investasi yang efisien [1].

Pada tahun 2023, Saenz dkk melakukan penelitian menggunakan data dari 240 perusahaan yang terdaftar di indeks *Russell 3000*, yang mencakup periode dari tahun 2017 hingga 2022. Dalam studi ini, saham-saham tersebut dikelompokkan menggunakan teknik K-Means dan berbagai metrik jarak, termasuk *Dynamic Time Warping* (DTW), *Fourier decomposition*, dan *Extended Frobenius Norm* (EROS). Fokus dari pengelompokan ini adalah pada rasio keuangan, harga, dan *return* harian. Model prediksi yang digunakan adalah LSTM. Model LSTM ini dilatih untuk masing-masing *cluster* dengan tujuan memprediksi harga harian, yang selanjutnya diubah menjadi *return* saham. Hasil penelitian menunjukkan bahwa model LSTM secara signifikan unggul

dalam memprediksi *return* saham, dengan rata-rata keuntungan *return* sebesar 1.97% dan standar deviasi sebesar 2.01%. Keberhasilan ini menegaskan bahwa penggunaan LSTM tidak hanya memberikan akurasi yang tinggi tetapi juga dalam performa, dan pengelompokan data *cluster* dapat meningkatkan keakuratan prediksi [4].

Dalam tugas akhir ini, *clustering* saham akan dilakukan dengan metode K-Medoids untuk mengelompokkan saham berdasarkan rata-rata *return* dan standar deviasi *return* untuk setiap saham. Selanjutnya, akan dilakukan prediksi harga saham yang diturunkan menjadi prediksi *return* menggunakan metode LSTM pada setiap *cluster*. Kinerja prediksi dari setiap *cluster* akan diukur menggunakan RMSE. Saham dengan hasil prediksi *return* tertinggi dari setiap *cluster* akan dipilih untuk mewakili *cluster* tersebut dalam pembentukan portofolio, proses seleksi ini dilakukan berdasarkan hasil prediksi *return* harian untuk setiap saham. Portofolio ini akan dibentuk menggunakan metode Equal Weight dan diukur menggunakan rata-rata *return* portofolio dan standar deviasi *return* portofolio akan dibandingkan dengan portofolio indeks *LQ45*.

#### **Topik dan Batasannya**

Dalam penelitian ini, penulis membangun model prediksi *return* menggunakan metode LSTM untuk setiap kelompok saham yang telah dikelompokkan menggunakan K-Medoids berdasarkan rata-rata *return* dan standar deviasi *return*. Selanjutnya, penulis membentuk portofolio saham yang mempertimbangkan saham dengan prediksi *return* tertinggi dari setiap kelompok, menggunakan metode Equal Weight. Kinerja portofolio diukur menggunakan rata-rata *return* dan standar deviasi *return*, dan hasilnya dibandingkan dengan indeks *LQ45*.

Batasan masalah pada penelitian ini adalah harga saham penutupan (*close*) dari data harian (*daily*) untuk 39 data saham dalam indeks *LQ45* yang didapatkan dari [finance.yahoo.com](https://finance.yahoo.com), dalam kurun waktu selama selama 10 tahun terakhir (01 Januari 2013 – 01 Januari 2023).

#### **Tujuan**

Penelitian ini bertujuan untuk membangun model prediksi yang dapat diandalkan dalam memprediksi harga saham yang dikonversikan menjadi *return* berdasarkan *clustering*. Keberhasilan model akan diukur menggunakan RMSE untuk mengukur kesalahan prediksi model. Tahapan penelitian meliputi pengelompokan saham menggunakan metode K-Medoids berdasarkan rata-rata *return* dan standar deviasi *return*, serta penerapan model LSTM untuk memprediksi *return* saham. Selain itu, penelitian ini juga bertujuan untuk menerapkan metode ini dalam proses seleksi saham untuk pembentukan portofolio yang terdiri dari saham dengan prediksi *return* tertinggi dari setiap *cluster*, yang selanjutnya pembentukan portofolio *equal weight* tersebut dibandingkan dengan kinerja indeks *LQ45*.

#### **Organisasi Tulisan**

Organisasi tulisan pada penelitian ini setelah pendahuluan yaitu berupa studi terkait mengenai penelitian terkait, K-Medoids *clustering*, metode model LSTM. Pada bagian sistem yang dibangun berisikan alur kerja sistem secara umum dari awal sampai akhir. Bagian selanjutnya yaitu berupa evaluasi yang berisikan hasil pengujian prediksi *return* untuk setiap *cluster*, pertumbuhan portofolio, dan perbandingan portofolio *equal weight* dengan indeks *LQ45*. Bagian terakhir berisikan kesimpulan pada penelitian ini.