1. Pendahuluan

1.1. Latar Belakang

Pada era digital dimana informasi dan data dapat dibagikan dan disebarluaskan dengan mudah, berbagai informasi dan data yang bersifat umum maupun pribadi, seperti nomor kartu kredit, perbankan, riwayat kesehatan dan informasi pribadi lainnya mulai digitalisasikan [1][2]. Data – data digital tersebut disimpan dan dibagikan melalui perangkat-perangkat pribadi seperti smartphone, laptop dan komputer, sehingga lebih mudah untuk disebarluaskan dan meminimalisir waktu serta biaya yang dibutuhkan dalam membagikan infromasi maupun data tersebut. Namun dibalik kemudahan yang ditawarkan, digitalisasi data dan informasi juga membawa masalah baru, dimana data - data atau informasi yang telah di digitalisasikan dapat dicuri atau digunakan untuk melakukan tindak kejahatan dan hal - hal tidak bertanggung jawab. Salah satu contohnya adalah *identity theft* (pencurian identitas), dimana pada tahun 2018, terdapat 14,4 juta orang yang menjadi korban pencurian identitas dan sekitar 80% perusahaan - perusahaan besar di dunia menderita kerugian akibat pencurian identitas [3]. Sony selaku salah satu perusahaan raksasa yang berasal dari jepang melaporkan kerugian senilai $100 juta dolar akibat pencurian identitas [3]. Selain *identity theft*, data atau informasi pribadi yang digitalisasikan juga dapat menjadi acuan tindak kejahatan seperti penculikan, pembegalan dan pemerasan. Namun, dengan adanya ancaman - ancaman tersebut, langkah - langkah atau pendekatan yang dilakukan dalam melindungi data - data digital masih terbilang lemah [2][4]. Penggunaan *password*, *single sign on* (SSO) dan *one time password* (OTP) sebagai langkah pengamanan data sudah menjadi hal umum dalam beberapa tahun terakhir dan pendekatan - pendekatan tersebut tidak sekuat seperti apa yang terlihat [1][2]. Pendekatan - pendekatan tersebut rentan terhadap *hackers* maupun penguping/pengintip, dimana *hackers* dapat meretas perangkat atau memecahkan *password* yang digunakan pemilik perangkat dan penguping/pengintip dapat menguping/mengintip pemilik perangkat ketika akan mengakses perangkatnya menggunakan *password* miliknya [2][4]. Oleh karena itu, untuk mengatasi hal tersebut banyak peneliti yang mulai mencari alternatif lain dalam mengamankan data dan informasi digital, dimana banyak peneliti yang menyarankan penggunaan *behavioral biometrics* sebagai sistem pengamanan [2][4][7].

Behavioral biometrics, khususnya *dynamic keystroke* memiliki beberapa keunggulan yang signifikan dibandingkan *physical biometric*, dimana pengaplikasian *dynamic keystroke* tidak memerlukan biaya yang besar, karena hanya memerlukan perangkat keras berupa *keyboard* dan penggunaannya yang tidak mengganggu aktifitas pengguna, karena data identifikasi dapat diambil bahkan tanpa pengguna sadari [1][3][5]. Terdapat tiga pendekatan utama yang digunakan dalam studi *dynamic keystroke*, yaitu *distance based*, *statistical based* dan *machine learning based*, dimana pendekatan yang sering digunakan adalah *distance based* dan *machine learning based* [3][4][5]. Penelitian - penelitian sebelumnya terkait *dynamic keystroke* telah menunjukan potensi *dynamic keystroke* dalam meningkatkan sekuritas sistem keamanan data pengguna, dimana *dynamic keystroke* menggunakan pola ketikan pengguna untuk mengidentifikasi individu sebagai pengguna yang sah atau penyusup, sehingga walaupun seorang penyusup dapat mengetahui informasi keamanan milik korban, seperti *password, single sign in on* (SSO) dan *one time password* (OTP), penyusup tidak dapat mengakses perangkat atau akun korban, karena pola pengetikan merupakan perilaku unik setiap individu yang sulit untuk direplikasi [2][4].

Pada tahun 2020 Lo et al. melakukan penelitian terkait *dynamic keystroke* dengan menggunakan pendekatan *distance based* dan *machine learning based*. Mereka menemukan bahwa *dynamic keystroke* dengan pendekatan *machine learning*, khususnya algoritma *Random Forest*, mengungguli pendekatan *distance based* dengan akurasi yang mencapai 95% [3]. Sama halnya dengan penelitian Lo et al, penelitian - penelitian yang dilakukan oleh beberapa peneliti seperti Graham et al. (2019), Darabseh et al. (2020), Pirzado et al. (2021) dan Chang et al. (2022) menunjukan bahwa identifikasi individu menggunakan *dynamic keystroke* dengan algoritma *Random Forest* memiliki akurasi yang cukup tinggi, dimana akurasi yang dicapai berada diatas 90% [4][5][6][7].

Meskipun penelitian - penelitian tersebut telah menunjukan hasil yang menjanjikan [3][4][5][6][7]. Sebagian besar penelitian tersebut dilakukan pada *dataset* yang berbasis *fixed text*, dimana berdasarkan penelitian yang dilakukan oleh Sim et al. penggunaan satu *N-graph*, seperti *digraph* dan *trigraph* dalam mengklasifikasikan kelas individu, kurang diskriminatif untuk *dataset* berbasis *free text* [19]. Pendekatan ini memiliki keterbatasan dalam menangkap pola yang kompleks, karena setiap *N-graph* hanya mampu menangkap aspek tertentu dari pola pengetikan suatu individu[19]. Sehingga untuk mengatasi hal tersebut, metode *multimodal N-graph* diterapkan dalam melakukan klasifikasi, dimana metode ini menggabungkan beberapa jenis fitur *N- graph (monograph, digraph* dan *trigraph)* untuk menghasilkan fitur fitur yang lebih kaya dan komprenhensif dalam mengidentifikasi identitas suatu individu.

Dengan mempertimbangkan penelitian - penelitian sebelumnya. Diperlukan penelitian lebih lanjut terkait kinerja *dynamic keystroke* dengan menggunakan algoritma *Random Forest*. Maka dari itu dilakukan penelitian terkait

multimodal N-graph dynamic keystroke, dimana penelitian ini akan menggunakan tiga buah *N-graph (monograph, digraph* dan *trigraph)* dan dua buah *dataset* (Biomey dan Alto) dengan algoritma *Random Forest*.

1.2. Topik dan Batasannya

Berdasarkan latar belakang yang telah dibahas sebelumnya, berikut adalah rumusan masalah yang akan dibahas pada penelitian ini :

1. Bagaimana cara menerapkan konsep *multimodal N-graph* terhadap *dynamic keystroke ?*
2. Bagaimana performansi algoritma *Random Forest* dalam melakukan klasifikasi pada *dynamic keystroke* dengan pendekatan *multimodal N-graph* ?
3. Bagaimana performansi *dynamic keystroke* menggunakan algoritma *Random Forest* dalam melakukan klasifikasi individu pada berbagai *N-graph* berbeda (*monograh, digraph* dan *trigraph*) ?

Agar penelitian tidak melebar, batasan masalah pada penelitian ini adalah :

1. *Dataset* yang digunakan pada penelitian ini adalah *dataset* Biomey dan Alto.
2. Sistem *keystroke* yang dibangun hanya mencakup *dataset* yang berbasis *free text.*
3. N-graph yang digunakan pada penelitian ini mencakup *monograph, digraph* dan *trigraph*.

1.3. Tujuan

Adapun tujuan dari penelitian ini adalah untuk mengembangkan teknologi *dynamic keystroke* yang lebih handal dan meneliti kinerja *dynamic keystroke*, yang mencakup :

1. Menerapkan konsep *multimodal N-graph pada dynamic keystroke*.
2. Meneliti performansi algoritma *Random Forest* dalam melakukan klasifikasi pada *dynamic keystroke* jika menggunakan pendekatan *multimodal N-graph.*
3. Meneliti performansi *dynamic keystroke* menggunakan algoritma *Random Forest* dalam melakukan klasifikasi individu pada berbagai *N-graph* berbeda (*monograph, digraph* dan *trigraph*).

1.4. Organisasi Tulisan

Setelah memaparkan pendahulan pada sub bab pertama, sub bab kedua pada jurnal ini akan membahas kajian kajian terkait *keystroke biometric* dan algoritma pembelajaran mesin *Random Forest*, sedangkan sub bab ketiga akan berfokus pada sistem yang dikembangkan serta *dataset* yang digunakan dalam penelitian. Sub bab ke empat akan membahas hasil evaluasi penelitian dan pengujian yang telah dilakukan. Pada sub bab terakhir yaitu sub bab kelima akan disajikan kesimpulan dan saran, yang dihasilkan dari hasil evaluasi penelitian yang telah dilakukan.