BAB I
PENDAHULUAN

1.1 Latar Belakang dan Identifikasi Masalah

Masalah klasifikasi sering dijumpai dalam kehidupan sehari – hari. Baik itu pengklasifikasian data pada bidang akademik, sosial, pemerintahan, maupun pada bidang lainnya. Klasifikasi merupakan suatu cara untuk mengelompokkan sesuatu sesuai dengan ciri – ciri tertentu. Klasifikasi sebagai aktivitas ilmiah yang dapat kita gunakan untuk mendeskripsikan, meringkas dan menyederhanakan suatu data ke dalam suatu format yang diinginkan sehingga memudahkan kita dalam melakukan penelitian. Klasifikasi merupakan proses kegiatan basis data yang fungsinya untuk menemukan properti - properti yang sama dalam himpunan obyek pada sebuah basis data kemudian basis data tersebut terklasifikasi menjadi beberapa kelas yang berbeda menurut model klasifikasi tertentu (Wismarini,et al., 2014).

Klasifikasi dalam ilmu statistik memiliki dua makna yang berbeda. Pertama, jika memiliki data yang sangat banyak sehingga perlu dilakukan peringkasan data dengan mengelompokkan data agar mempermudah dalam analisis dan penggunaan data. Kedua, jika mempunyai banyak kategori atau kelompok, sehingga perlu membuat klasifikasi atau kelompok yang baru yang bisa mencakup banyak kelompok tadi. (Michie,et al., 1994).

Klasifikasi dalam ilmu statistik digunakan untuk mengelompokkan suatu data yang disusun secara sistematis dengan mengasumsikan sebagai banyaknya

kategori dari suatu individu dan setiap populasi dikarakteristikkan dengan ukuran distribusi probabilitasnya. Teknik klasifikasi dalam statistik terbagi menjadi dua bagian yaitu teknik klasik (classical statistic model) dan teknik modern (modern statistic model). Teknik klasifikasi klasik memiliki beberapa peraturan atau asumsi yang perlu untuk dipenuhi sedangkan Teknik klasifikasi modern lebih fleksibel, tidak terikat dengan aturan atau asumsi yang berlaku. Teknik klasifikasi statistik yang klasik terdiri dari analisis diskriminan linier, diskriminan kuadratik, diskriminan logistik dan regresi logistik. Sedangkan untuk teknik klasifikasi modern terdiri dari KNN (K-Nearest Neighbour), naive bayes, jaringan saraf, dan klasifikasi projeksion pursuit (projection pursuit classification) (Michie, et al., 1994).

Analisis diskriminan merupakan salah satu metode statistik untuk klasifikasi suatu data. Analisis diskriminan adalah analisis klasifikasi yang biasa dan sering kita dengar setelah analisis regresi logistik. Analisis diskriminan merupakan metode statistik yang digunakan untuk mengklasifikasikan suatu individu atau observasi ke dalam suatu kelas atau kelompok berdasarkan sekumpulan variabel (Johnson dan Wichern, 2007). Tujuan utama analisis diskriminan untuk mengetahui ada atau tidaknya perbedaan antar suatu kelompok (Hair J, Black, dan Anderson, 2010).

Analisis diskriminan sering digunakan untuk penelitian perusahaan agar perusahaan bisa berkembang pesat. Sebagai contoh yaitu penelitian dari Ramadhan et al., (2018) yang membahas tentang keputusan masyarakat untuk membeli produk McCafe di McDonald Jimbaran Bali menggunakan analisis diskriminan. Penelitian

tersebut dilatarbelakangi persaingan perusahaan makanan yang semakin ketat sehingga perlu dilakukan penelitian untuk meningkatkan kualitas dari perusahaan. Namun saat ini penggunaan analisis diskriminan tidak hanya untuk perusahaan yang ingin berkembang namun juga untuk bidang yang lainnya seperti kependudukan dan kesehatan. Penelitian di bidang kependudukan juga ada yang menggunakan analisis diskriminan seperti penelitian dari Pratiwi, Rahardjo dan Susiswo (2012) mengenai analisis kelompok dan diskriminan untuk menggolongkan tingkat pengangguran di Jawa timur.

Selain Analisis Diskriminan, saat ini ilmu statistika telah mengembangkan teknik klasifikasi modern, salah satunya yaitu *naive bayes*. Teknik klasifikasi *naive bayes* merupakan salah satu dari teknik klasifikasi statistik modern dengan memanfaatkan probabilitas dan statistik. Metode ini dikemukakan oleh ilmuwan Inggris, *Thomas Bayes*. Tujuan metode ini untuk memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya. Teknik klasifikasi *naive bayess* memiliki keunggulan yaitu rumus sederhana dan memiliki tingkat akurasi yang cukup tinggi.

Friedman et al., (1997) menjelaskan bahwa pendekatan klasifikasi *bayesian network* yang dikenal paling efektif adalah *Naive Bayes* (NB). Pendekatan klasifikasi *Naive Bayes* merupakan pendekatan yang paling sederhana dengan menghubungkan semua variabel independen dengan variabel dependen. Asumsi dari pendekatan klasifikasi *Naive Bayes* adalah semua variabel harus saling independen.

Penelitian dari Afeni, B.O, Thomas I.A dan Iyanuoluwa A.O (2017) menghasilkan bahwa *Naive Bayes* terbukti memiliki tingkat ketepatan klasifikasi lebih tinggi dibandingkan dengan *decision tree*. Hal ini juga didukung oleh penelitian dari Iyer, Jeyalatha dan Ronak S (2018) yang meneliti tentang diagnosa diabetes menggunakan teknik klasifikasi data mining. Penelitian tersebut membandingkan dua teknik klasifikasi yaitu J48 pohon keputusan dengan *Naive Bayes.* Hasil yang didapat adalah dari kedua metode yang dibandingkan, metode *Naive Bayes* terbukti memiliki ketepatan klasifikasi lebih tinggi dibandingkan dengan J48 pohon keputusan.

Kedua teknik klasifikasi diatas memiliki perbedaan, Analisis Diskriminan merupakan teknik klasifikasi klasik sedangan *Naive Bayes* adalah teknik klasifikasi modern yang menggabungkan teknik statistika dengan data mining dan memiliki kemampuan mengklasifikasikan dengan nilai akurasi yang tinggi dikarenakan *naive bayes* sangat cocok digunakan pada data skala besar dan baik untuk data bias. Namun, belum ada penelitian yang membandingkan kedua teknik klasifikasi tersebut terutama ketika variabel yang dimiliki berupa data kategori dan kontinyu. Penelitian ini mengambil data dari hasil Survei Demografi dan Kesehatan Indonesia (SDKI) tahun 2017 dari kasus Preferensi Fertilitas di Provinsi Jawa Timur.

Hasil sensus penduduk pada tahun 1980, 1990, 2000 dan 2010 berturut turut menunjukkan jumlah penduduk di Provinsi Jawa Timur sebesar 29.188.852 jiwa, 32.503.815 jiwa, 34.765.998 jiwa dan 37.4776.757 jiwa. Hasil sensus penduduk tersebut menunjukkan bahwa jumlah penduduk di Provinsi Jawa Timur menempati urutan ke 2 terbanyak setelah provinsi Jawa Barat. Provinsi Jawa Timur memiliki

laju pertumbuhan penduduk selama periode 1980-1990, 1990-2000 dan 2000-2010 berturut turut sebesar 1,08 persen per tahum, 0,70 persen per tahun dan 0,76 persen per tahun (BPS, 2018).

Jumlah penduduk di Provinsi Jawa Timur diprediksi akan semakin bertambah dari tahun ke tahun. Jumlah penduduk di jawa Timur tahun 2015 mencapai 38.847.561 jiwa dan akan semakin naik menjadi 39.698.631 jiwa pada tahun 2019. Pertumbuhan penduduk akan menjadi modal pembangunan sebuah negara. Semakin banyak jumlah penduduk, kebutuhan akan sandang, pangan, papan, pendidikan dan lapangan kerja akan meningkat sehingga harus menjadi perhatian dari pemerintah.

Pertambahan jumlah penduduk tidak terlepas dari tingkat fertilitas (kelahiran). Tingginya fertilitas akan menimbulkan berbagai masalah sosial terlebih jika tidak dibarengi dengan peningkatan kualitas sumber daya manusia. Tingginya fertilitas dapat mencerminkan rata-rata usia kawin rendah atau banyaknya usia pernikahan dini, tingkat pendidikan rendah serta tingkat sosio ekonomi juga rendah. Penjelasan tersebut memberi kesimpulan bahwa perlu dilakukan berbagai upaya untuk menekankan tingkat fertilitas.

1.2 Kajian Masalah

Analisis Diskriminan merupakan salah satu teknik di bidang ilmu statistika yang digunakan untuk klasifikasi individu atau objek ke dalam kelompok yang saling bebas dan menyeluruh berdasarkan kumpulan dari peubah atau diskriminatornya. Analisis Diskriminan memiliki manfaat sebagai alat untuk

memisahkan segugus pengamatan dan mengklasifikasikan pengamatan baru ke dalam kelompok yang sudah didefinisikan sebelumnya. Terdapat dua asumsi utama yang harus dipenuhi pada analisis diskriminan, yaitu sejumlah diskriminator atau peubah penjelas harus memiliki distribusi normal dan matriks ragam beragam dari diskriminator atau peubah penjelas harus sama dan skala data yang dipakai adalah variabel yang memiliki skala data kontinu.

Namun, dalam prakteknya sering kali dalam banyak kasus tidak semua peubah – peubah atau variabel prediktor berjenis kontinyu, beberapa diantaranya berjenis data kategorik. Hal ini ditunjukkan dari penelitian oleh Walidaini, dkk (2017) yang berjudul analisis diskriminan berganda dengan peubah bebas campuran kategorik dan kontinu pada klasifikasi indeks prestasi kumulatif mahasiswa. Pada penelitian tersebut, variabel yang dipakai variabel kontinu dan kategorik terutama untuk variabel jenis kelamin.

Teknik klasifikasi modern *Naive Bayes* memiliki rumus yang sederhana dan tingkat akurasi yang tinggi. Hal ini ditunjukkan dari penelitian Kabir F et al., (2011) menunjukkan bahwa hasi dari ketepatan klasifikasi *Naive Bayes* cukup tinggi sebesar 94.14%. Redjeki Sri, M.Guntara dan Pius A (2015) juga menjelaskan dalam penelitiannya mengenai klasifikasi masyarakat miskin di kabupaten Bantul menggunakan *Naive Bayes* memiliki tingkat ketepatan klasifikasi sebesar 93.18%. Vembandasamy K, et al., (2015) melakukan diagnosa penyakit jantung menggunakan *Naive Bayes* dan dihasilkan ketepatan klasifikasi benar sebesar 86.41% dan tidak benar sebesar 13.58%. secara keseluruhan, ketepatan klasifikasi *Naive Bayes* pada penelitian diagnosa penyakit jantung sebesar 74%.

Penelitian terdahulu menyebutkan bahwa kedua metode klasifikasi (Analisis Diskriminan dan *Naive Bayes*) memiliki tingkat akurasi untuk klasifikasi cenderung tinggi. Perbandingan antara Analisis Diskriminan dengan *Naive Bayes* sudah dilakukan namun belum ada penelitian yang membandingkan Analisis Diskriminan ketika data yang dipakai kategori dan numerik dengan metode klasifikasi modern *Naive Bayes* sehingga peneliti ingin membandingkan Analisis Diskriminan tersebut dengan *Naive Bayes*, apakah tetap konsisten menunjukkan tingkat akurasi yang tinggi serta dari kedua klasifikasi tersebut manakah yang memiliki akurasi klasifikasi yang tinggi. Penerapan kasus yang dipakai dalam penelitian ini adalah preferensi fertilitas.

Jumlah penduduk Provinsi Jawa Timur pada tahun 2017 sebesar 39.292.972 jiwa mengalami peningkatan dibandingkan tahun 2016 sebesar 39.075.152 jiwa (BPS Prov.Jatim, 2018). Salah satu faktor yang mempengaruhi laju pertumbuhan penduduk adalah tingkat fertilitas (kelahiran). Kondisi fertilitas di Indonesia sebenarnya telah menurun selama 10 tahun terakhir. Hasil dari Survei Demografi dan Kesehatan Indonesia (SDKI) tahun 2002/2003, 2007 dan 2012 menunjukkan bahwa angka Total Fertility Rate (TFR) di Indonesia berada di angka 2,6 anak sedangkan tujuan Rencana Pembangunan Jangka Menengah Nasional (RPJMN) menargetkan angka TFR sebesar 2,1 pada tahun 2015.

Angka TFR Provinsi Jawa Timur dari tahun 2012 sebesar 2,3 anak mengalami penurunan di tahun 2017 sebesar 2,1 anak. Namun, hasil dari SDKI 2017 mencatat bahwa angka fertilitas yang diinginkan Pasangan Usia Subur (PUS) mengalami peningkatan dari tahun 2012 sebesar 2,0 anak menjadi 2,1 di tahun

2017. Tingkat fertilitas ini sangat dipengaruhi oleh preferensi fertilitas tiap individu sehingga terdapat kemungkinan bahwa tingkat TFR disebabkan oleh preferensi fertilitas yang tidak sesuai dengan program BKKBN (Fadila W, 2015).

Studi preferensi fertilitas menentukan preferensi fertilitas seseorang secara berbeda. Pengukuran ini dapat disebut sebagai ukuran keluarga yang diinginkan, jumlah anak ideal, keinginan untuk anak tambahan, niat kesuburan dan sebagainya. Preferensi fertilitas merupakan ukuran penting untuk memperkirakan tingkat kelahiran yang diinginkan atau tidak diinginkan, memperkirakan kesuburan dan menilai kebutuhan kontrasepsi (Lunani, 2014).

Penelitian ini mempunyai tujuan untuk mengklasifikasikan preferensi fertilitas di Provinsi Jawa Timur. Penelitian mengenai klasifikasi preferensi fertilitas ini penting untuk dilakukan karena keputusan untuk memiliki anak lagi pada wanita kawin usia subur di Jawa Timur nantinya dapat berpengaruh terhadap penggunaan alat kontrasepsi. Ketika wanita kawin usia subur memutuskan untuk tidak memiliki anak lagi, maka mereka cenderung akan ber-KB atau memilih untuk memakai kontrasepsi (Kiswanto, Eddy., 2015).

Kinerja BKKBN Provinsi Jawa Timur telah dinilai baik, namun dengan adanya penelitian mengenai preferensi fertilitas wanita kawin usia subur di Provinsi Jawa Timur diharapkan dapat sebagai bahan pertimbangan tepat sasaran untuk program penyuluhan ber-KB dan evaluasi untuk BKKBN Provinsi Jawa Timur dalam menjalankan tugasnya sehingga capaian kinerja menjadi lebih baik pada tahun mendatang. Maka, penelitian ini bertujuan untuk mengklasifikasi preferensi

fertilitas pada Wanita Kawin Usia Subur di Provinsi Jawa Timur dengan menggunakan pendekatan Analisis Diskriminan dan *Naive Bayes* agar dapat menemukan pola faktor yang mempengaruhi preferensi fertilitas. Data yang dipakai dalam analisis adalah *row data* (data mentah) SDKI tahun 2017.

1.3 Batasan dan Rumusan Masalah

Berdasarkan sub bab sebelumnya, maka rumusan masalah pada penelitian ini adalah bagaimana klasifikasi preferensi fertilitas Wanita Kawin Usia Subur di Provinsi Jawa Timur dengan menggunakan pendekatan Analisis Diskriminan dan *Naive Bayes?*

1.4 Tujuan Penelitian

1.4.1 Tujuan Umum

Mengklasifikasikan preferensi fertilitas Wanita Kawin Usia Subur di Provinsi Jawa Timur dengan menggunakan pendekatan Analisis Diskriminan dan *Naive Bayes.*

1.4.2 Tujuan Khusus

1. Mengidentifikasi karakteristik wanita kawin usia subur di Provinsi Jawa Timur
2. Menganalisis variabel apa saja yang berperan penting dalam Analisis Diskriminan
3. Menganalisis tingkat ketepatan klasifikasi preferensi fertilitas wanita kawin usia subur dengan pendekatan Analisis Diskriminan

4. Menganalisis tingkat ketepatan klasifikasi preferensi fertilitas wanita kawin usia subur dengan pendekatan metode *Naive Bayes*
5. Membandingkan tingkat ketepatan klasifikasi dari pendekatan Analisis Diskriminan dan *Naive Bayes.*

1.5 Manfaat Penelitian

Penelitian ini diharapkan nantinya penggunaan metode Analisis Diskriminan dan *Naive Bayes* dapat menjadi alternatif kepada masyarakat, khususnya untuk menangani permasalahan dibidang kesehatan dan kependudukan. Selain itu juga penelitian ini mempunyai kegunaan sebagai berikut:

1. Bagi Instansi terkait

   Sebagai bahan masukan dan evaluasi untuk BKKBN Provinsi Jawa Timur dalam menjalankan tugasnya sehingga capaian kinerja menjadi lebih baik pada tahun mendatang.

2. Bagi Fakultas

   Sebagai tambahan ilmu pengetahuan dan kepentingan pembelajaran untuk mahasiswa lainnya.

3. Bagi Peneliti

   Sebagai media belajar bagi peneliti dalam menerapkan ilmu yang didapat selama pendidikan.