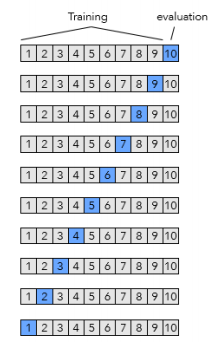
## **Skenario Evaluasi Pertama**

Skenario evaluasi pertama dalam penelitian ini dilakukan terhadap seluruh data yang berjumlah 700. Evaluasi ini dilakukan untuk menentukan perbandingan akurasi dari tiap percobaan nilai *k* pada K–*Nearest Neighbor* (KNN) dan *Modified* K–*Nearest Neighbor* (MKNN). Metode yang digunakan pada evaluasi ini adalah k–*fold cross–validation*.Dalamk–*fold cross–validation*,data secara acak akan dipartisi menjadi k–*fold* atau “lipatan” yang masing – masing berukuran sama. Sehingga, *training* dan *testing* akan dilakukan sebanyak k–*fold* kali. Dalam iterasi pertama, dipilih sebagai data *testing*,dan partisi yang tersisa yaitu ,….,secara kolektif digunakan sebagai data *training*. Sedangkan dalam iterasi kedua, partisi dipilih sebagai data *testing*, dan partisi , ,…,digunakan sebagai data *training*, dan begitu seterusnya sampai iterasi sebanyak k–*fold*.



Gambar 7.6. 10–Fold Cross–Validation

Untuk kasus klasifikasi menggunakan K–*Nearest Neighbor* (KNN) dan *Modified* K–*Nearest Neighbor* (MKNN), tiap data *training* dan data *testing* akan dihitung akurasinya dengan menggunakan *confusion matrix* dengan rumus yang dijelaskan pada persamaan (2.2). Untuk nilai k–*fold* yang digunakan pada metode ini adalah k = 10 sesuai dengan nilai 10–*fold* yang umum digunakan pada gambar 7.6. Sehingga, dari 700 yang digunakan akan dipartisi menjadi 10 bagian dengan masing – masing berisi 70 data. Pada evaluasi ini, diuji coba nilai *k* dari 1–50, dan untuk setiap nilai *k* akan dilakukan 10 kali iterasi, dimana partisi data yang digunakan sebagai data *training* dan data *testing* akan berbeda – beda di tiap iterasi sesuai dengan penjelasan diatas. Kemudian, untuk setiap kali iterasi akan dihitung jumlah yang benar dan tidak pada tabel data prediksi dan data aktual, sehingga dalam satu nilai *k* yang diuji coba dapat ditentukan akurasi nya sesuai dengan persamaan (2.2). Setelah semua nilai *k* diuji coba, maka didapatkan hasil akurasi dari tiap nilai *k* tersebut untuk kemudian ditentukan beberapa nilai *k* yang menghasilkan akurasi tertinggi.

## **Skenario Pengujian Kedua**

Pada skenario evaluasi kedua yang dilakukan terhadap seluruh 700 data, akan dihitung akurasinya menggunakan persamaan (2.2). Data evaluasi ini akan dilakukan seleksi fitur dengan *Principal Component Analysis* (PCA), dimana PCA akan mengurangi dimensi data dengan cara ‘mengkombinasikan’ intisari dari atribut dengan membentuk alternatif *subset* fitur yang lebih kecil menjadi terbentuk variabel fitur baru. Setelah itu, data PCA ini akan menentukan hasil klasifikasi dengan metode K–*Nearest Neighbor* (KNN) dan *Modified* K–*Nearest Neighbor* (MKNN) yang juga dilakukan dengan uji coba menggunakan rentang nilai *k* dari 1–50. Sehingga, setelah mendapatkan akurasi dari tiap percobaan nilai *k* pada skenario ini, akan ditentukan juga beberapa nilai *k* yang menghasilkan akurasi tinggi, kemudian beberapa nilai *k* dengan akurasi tertinggi pada skenario evaluasi pertama dan kedua akan diuji coba pada tahap pengujian.

## **Tahap Pengujian (Testing)**

Tahap Pengujian dalam penelitian ini dilakukan terhadap keseluruhan data yang berjumlah 700 data dan menggunakan metode k–*fold cross–validation* dengan nilai k = 10. Sehingga, dari 700 yang digunakan akan dipartisi menjadi 10 bagian dengan masing – masing berisi 70 data.

Pada tahap pengujian ini, diuji coba beberapa nilai *k* dengan akurasi tertinggi yang didapat dari skenario evaluasi pertama dan kedua. Kemudian, setiap nilai *k* yang diuji coba maka dapat dihitung akurasinya dengan *confusion matrix* dengan rumus yang dijelaskan pada persamaan (2.2) juga, dan berdasarkan hasil akurasi tersebut dapat ditentukan nilai *k* optimal yang menghasilkan akurasi tertinggi.