

K Nearest Neighbors

with

Iris Dataset

จัดทำโดย

นายศุภณัฐ แซ่เตีย

ID: 6505000270

Agenda

- รูปแบบในการทดสอบ
- การทดสอบ
 - ข้อมูลทรน 80% : ข้อมูลทดสอบ 20%
 - ข้อมูลทรน 70% : ข้อมูลทดสอบ 30%
 - ข้อมูลทรน 60% : ข้อมูลทดสอบ 40%

รูปแบบในการทดสอบ

- การแบ่งข้อมูล
 - ใช้ train test split ของ sklearn
 - กำหนดค่า random state ไว้ที่ 42 เพื่อให้การแบ่งข้อมูลในการทดลองในแต่ละครั้งมีผลเหมือนกัน
- การทรนโมเดล
 - ใช้ K Neighbors Classifier ของ sklearn
 - ทดลองใช้ค่า K ที่มีค่า Mean Error ต่ำที่สุด และสูงที่สุด
 - ทดลองใช้ weights
 - uniform : กำหนดให้เพื่อนบ้านทุกตัวมีน้ำหนักเท่ากัน
 - distance : กำหนดให้เพื่อนบ้านที่อยู่ใกล้จะมีน้ำหนักมากกว่า

คัดส่วนข้อมูล 80:20

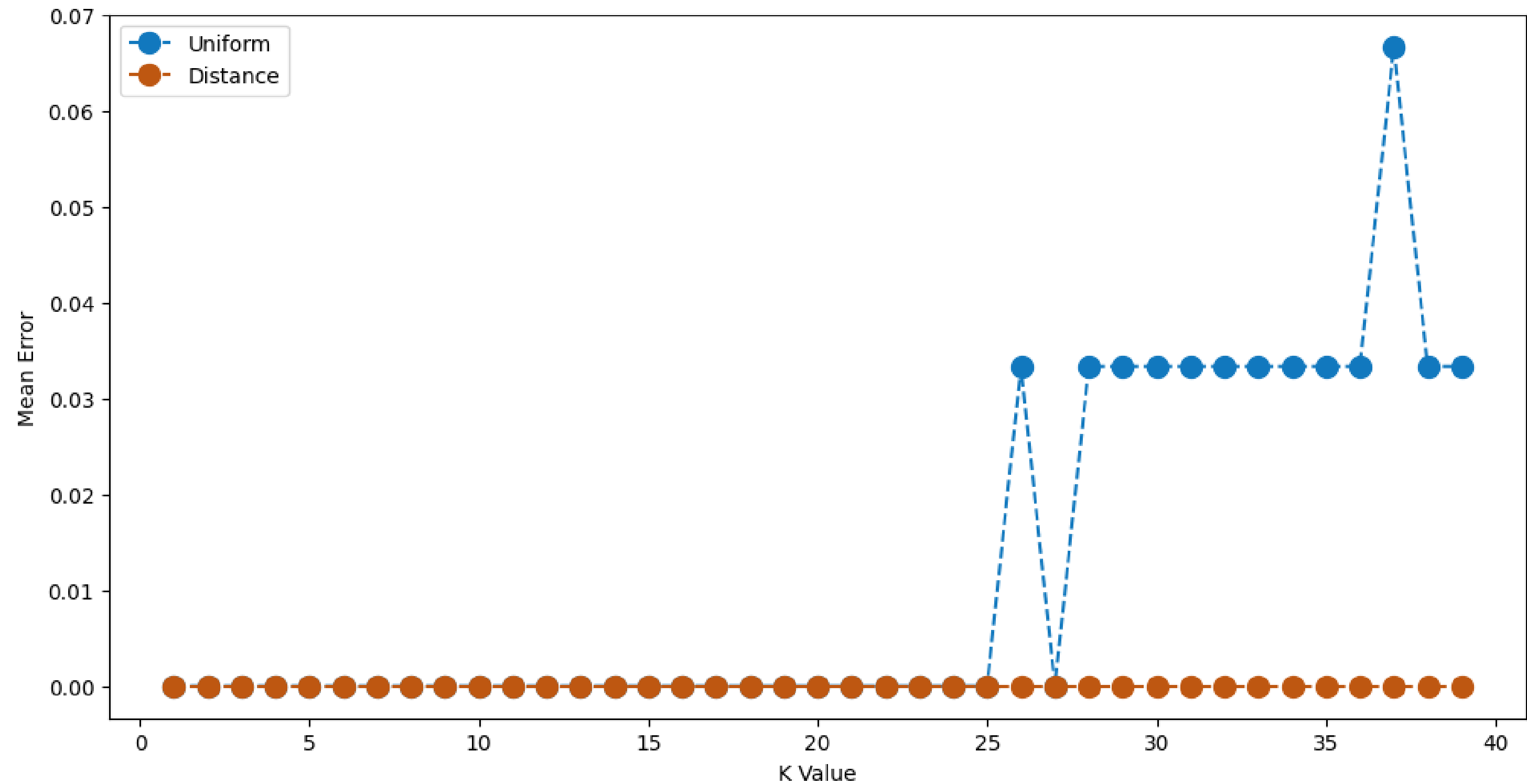
สัดส่วนข้อมูล 80:20

- แบ่งข้อมูลออกเป็น
 - ชุดข้อมูลสำหรับ Train : 120 ตัวอย่าง
 - ชุดข้อมูลสำหรับ Test : 30 ตัวอย่าง

	Train	Test
Iris-setosa	40	10
Iris-versicolor	41	9
Iris-virginica	39	11

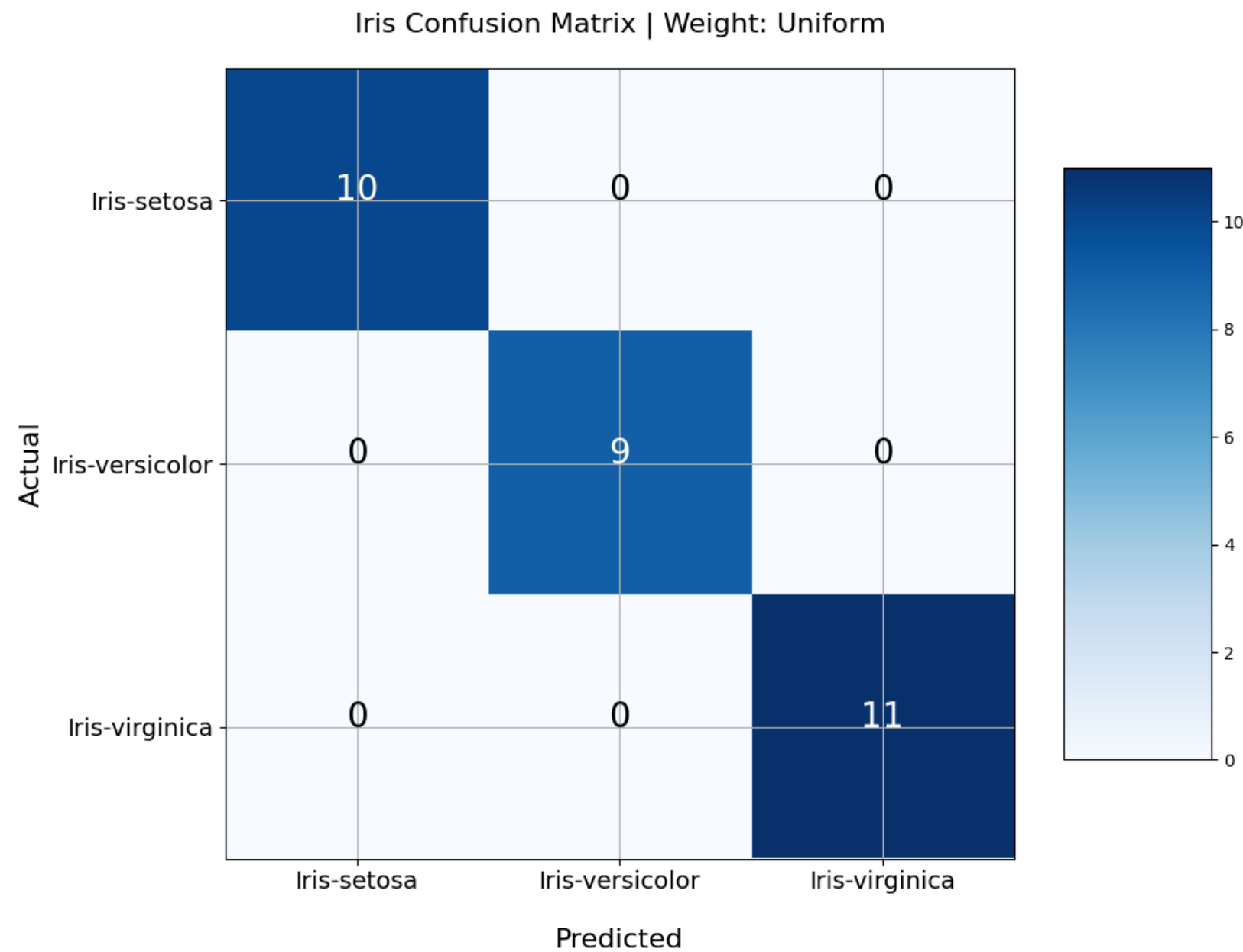
สัดส่วนข้อมูล 80:20

- Error Rate

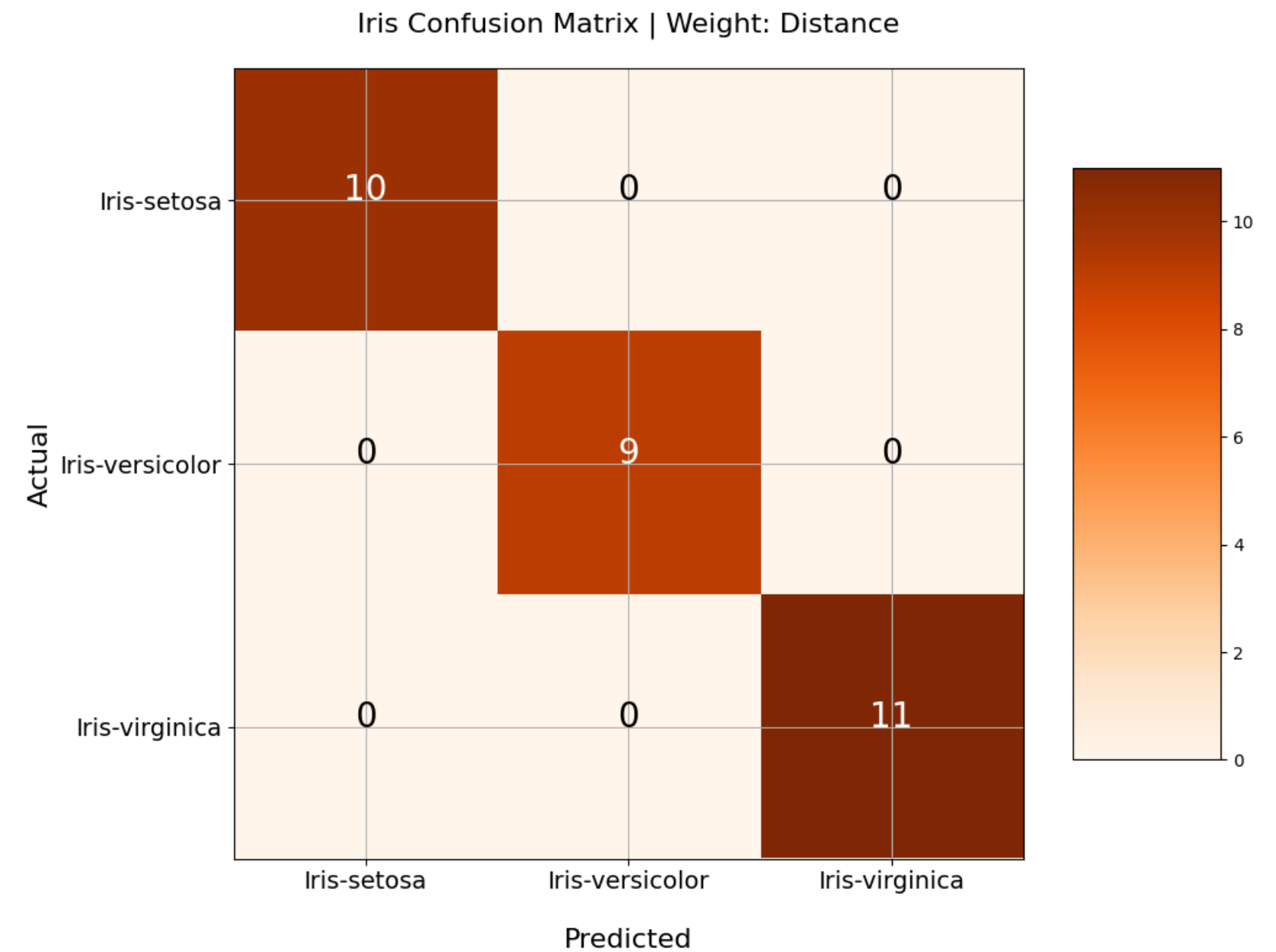


สัดส่วนข้อมูล 80:20

- ทดสอบด้วยค่า K ที่ Mean Error ต่ำสุด



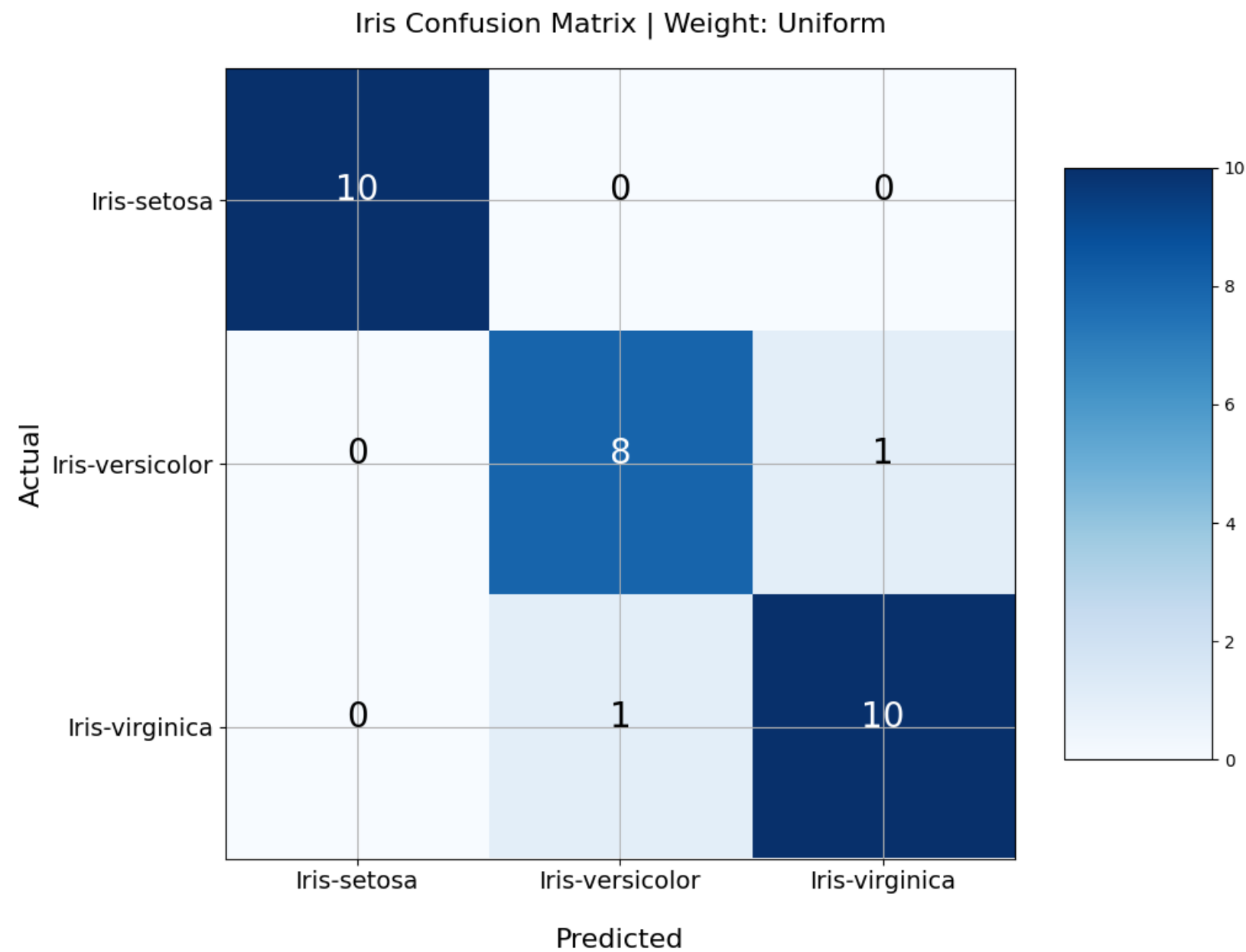
ความแม่นยำ: 100.00%



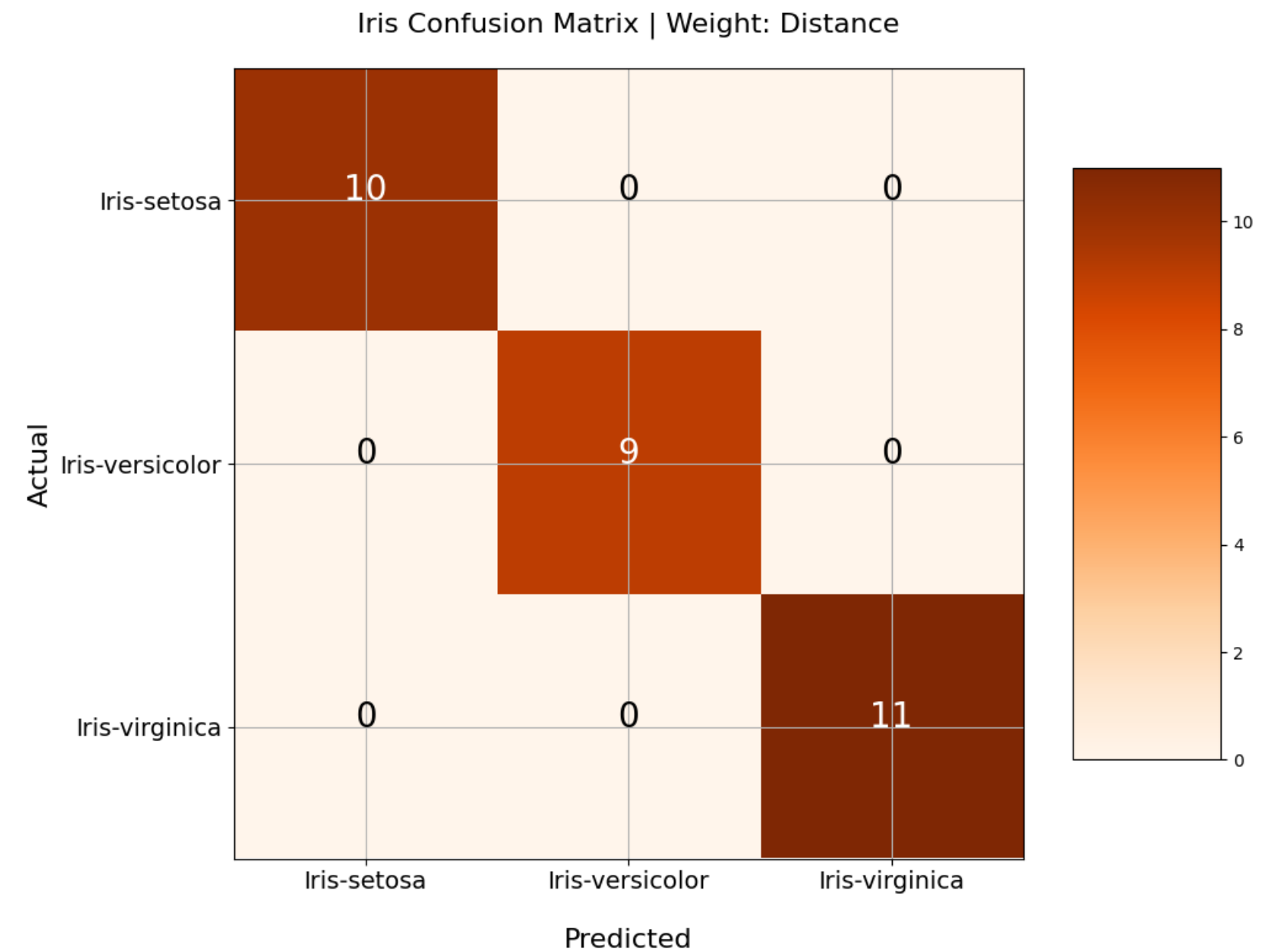
ความแม่นยำ: 100.00%

สัดส่วนข้อมูล 80:20

- ทดสอบด้วยค่า K ที่ Mean Error สูงสุด



ความแม่นยำ: 93.33%



ความแม่นยำ: 100.00%

ตัดส่วนข้อมูล 70:30

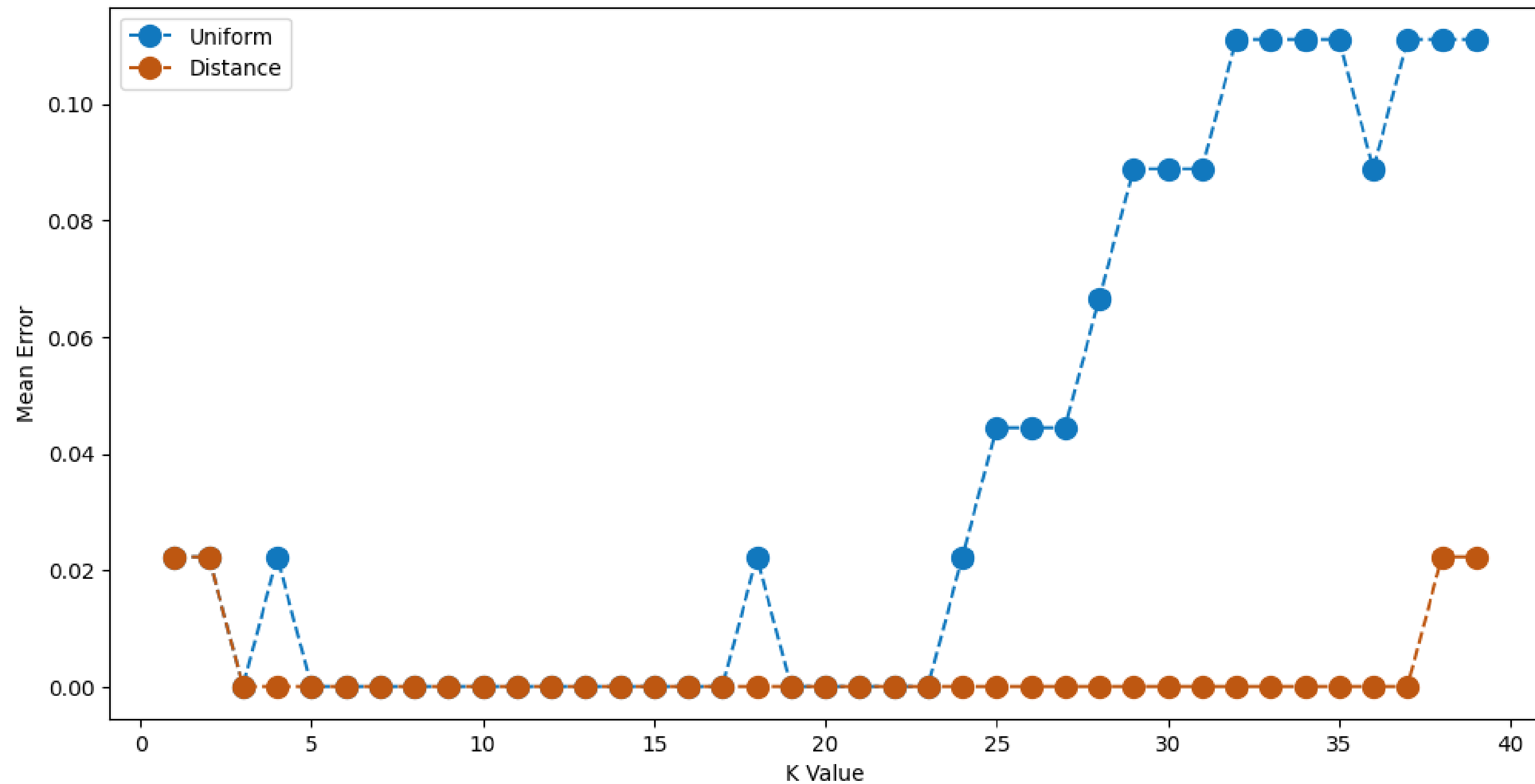
สัดส่วนข้อมูล 70:30

- แบ่งข้อมูลออกเป็น
 - ชุดข้อมูลสำหรับ Train : 105 ตัวอย่าง
 - ชุดข้อมูลสำหรับ Test : 45 ตัวอย่าง

	Train	Test
Iris-setosa	31	19
Iris-versicolor	37	13
Iris-virginica	37	13

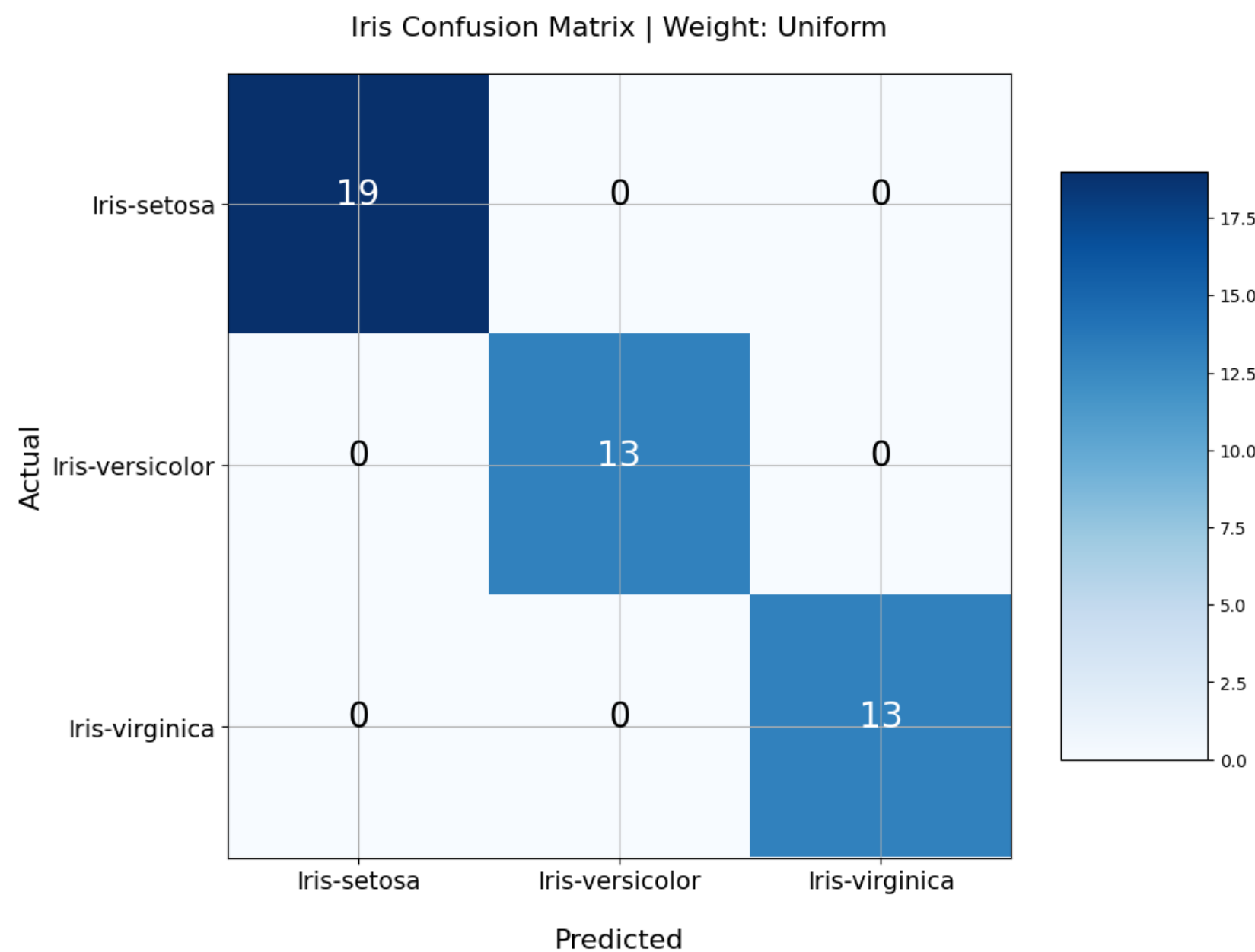
สัดส่วนข้อมูล 70:30

- Error Rate

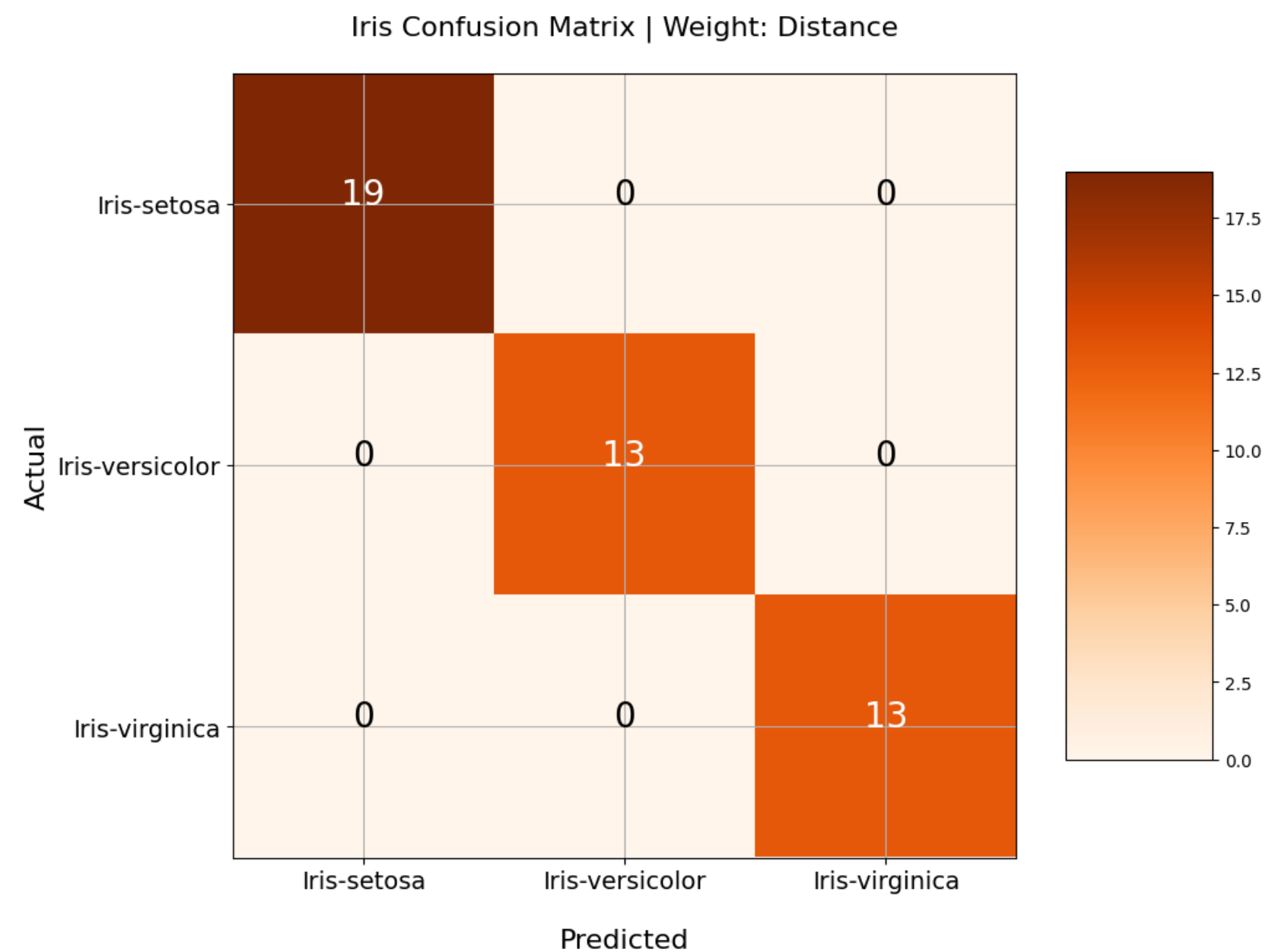


สัดส่วนข้อมูล 70:30

- ทดสอบด้วยค่า K ที่ Mean Error ต่ำสุด



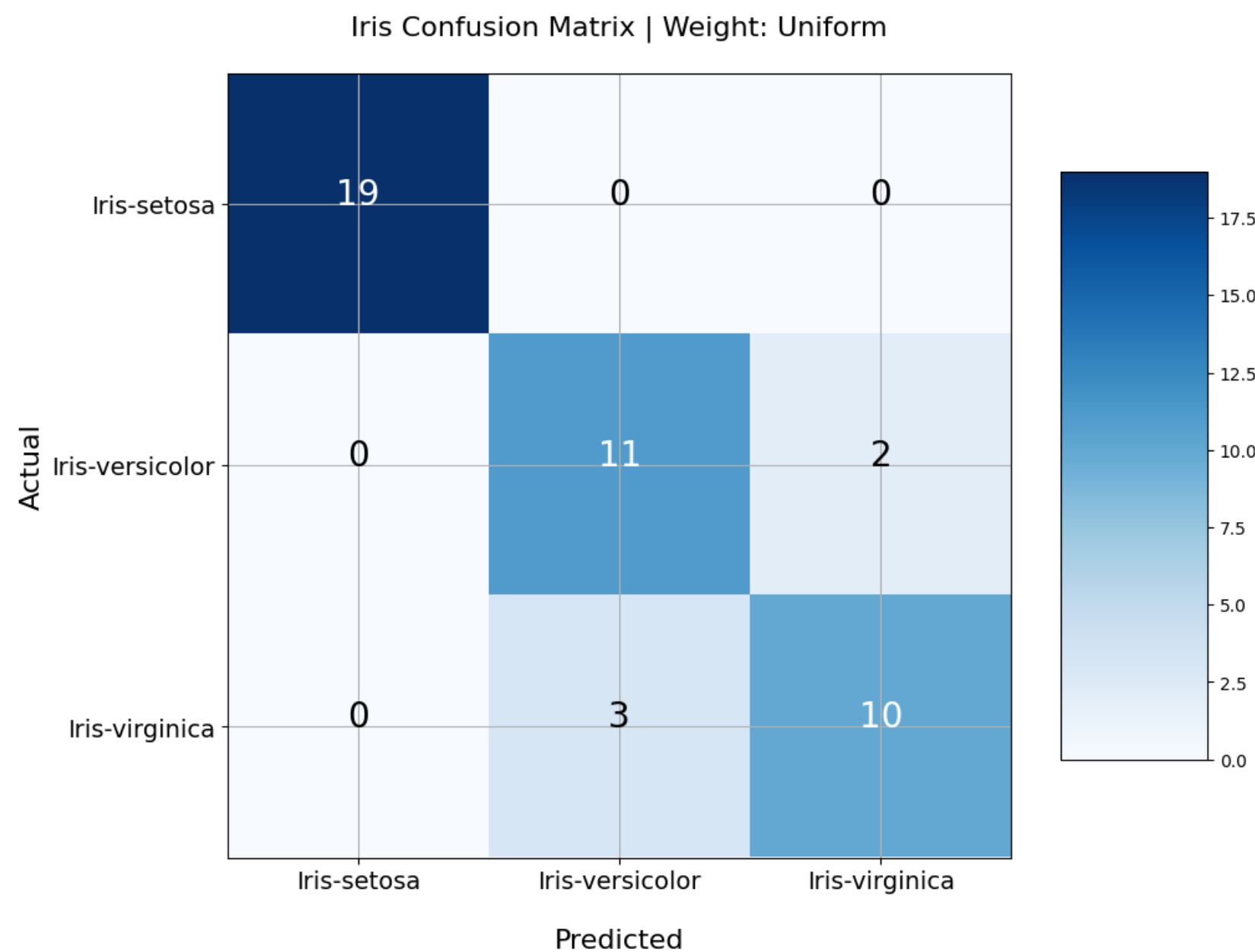
ความแม่นยำ: 100.00%



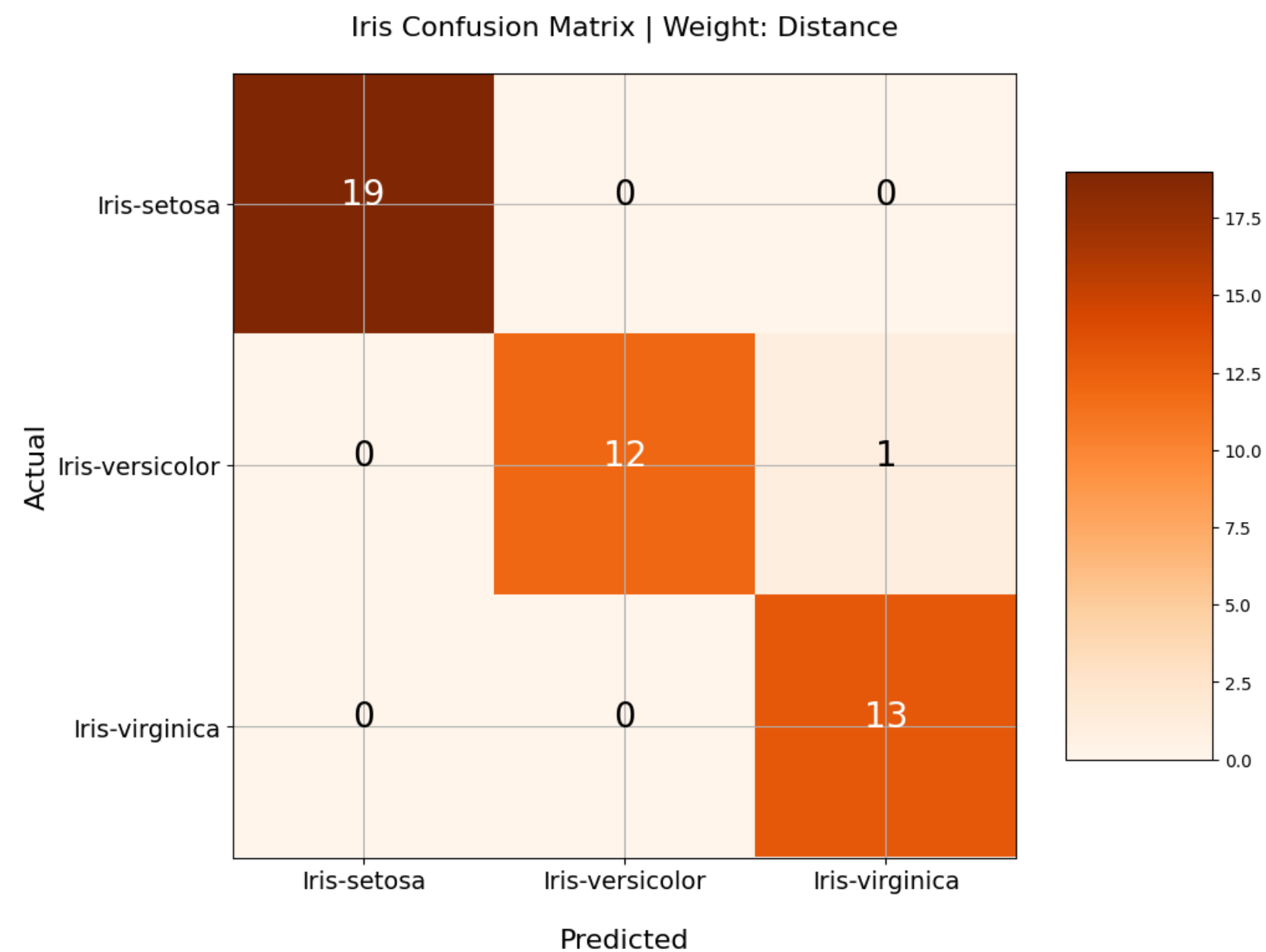
ความแม่นยำ: 100.00%

สัดส่วนข้อมูล 70:30

- ทดสอบด้วยค่า K ที่ Mean Error สูงสุด



ความแม่นยำ: 88.89%



ความแม่นยำ: 97.78%

ตัดส่วนข้อมูล 60:40

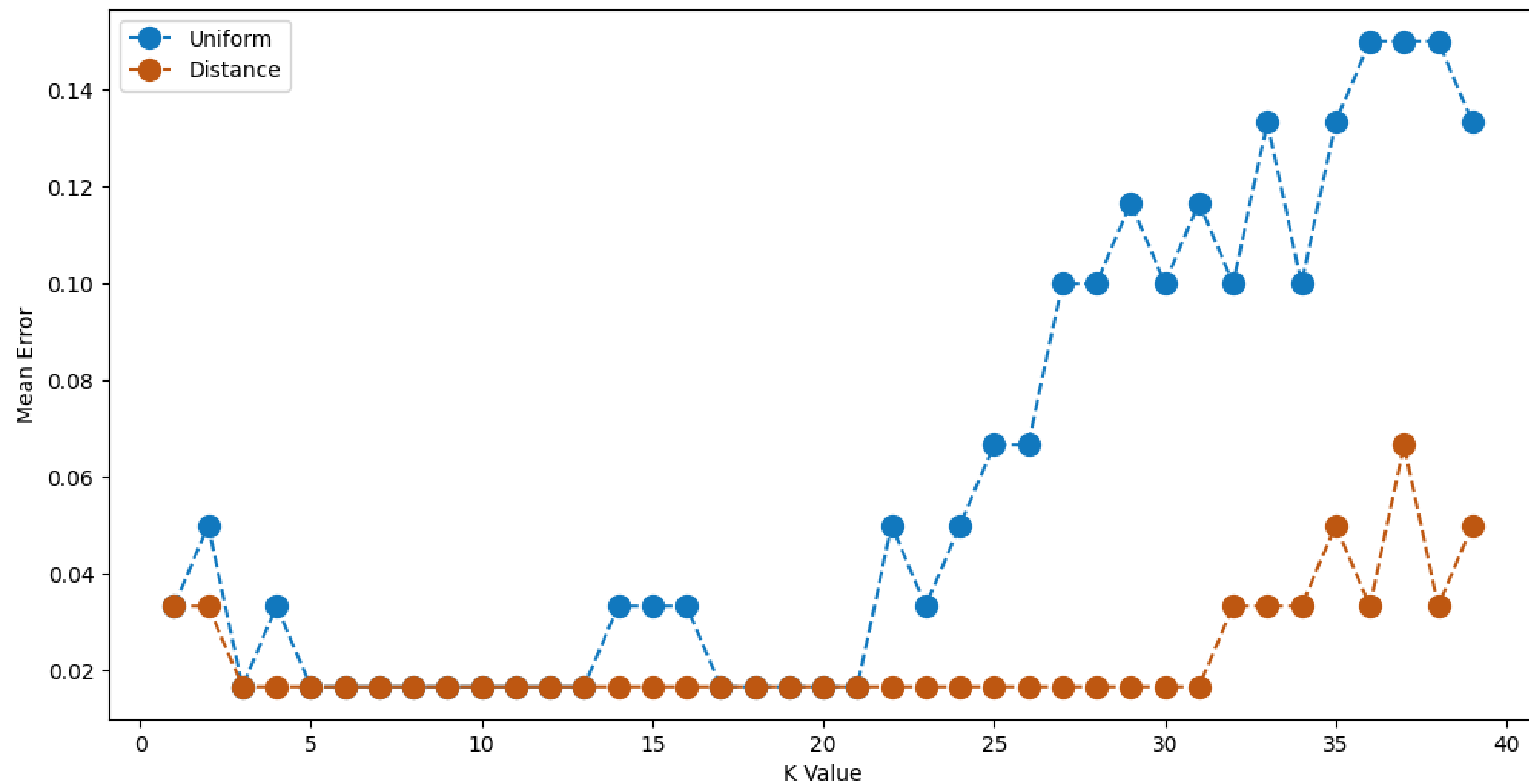
สัดส่วนข้อมูล 60:40

- แบ่งข้อมูลออกเป็น
 - ชุดข้อมูลสำหรับ Train : 90 ตัวอย่าง
 - ชุดข้อมูลสำหรับ Test : 60 ตัวอย่าง

	Train	Test
Iris-setosa	27	23
Iris-versicolor	31	19
Iris-virginica	32	18

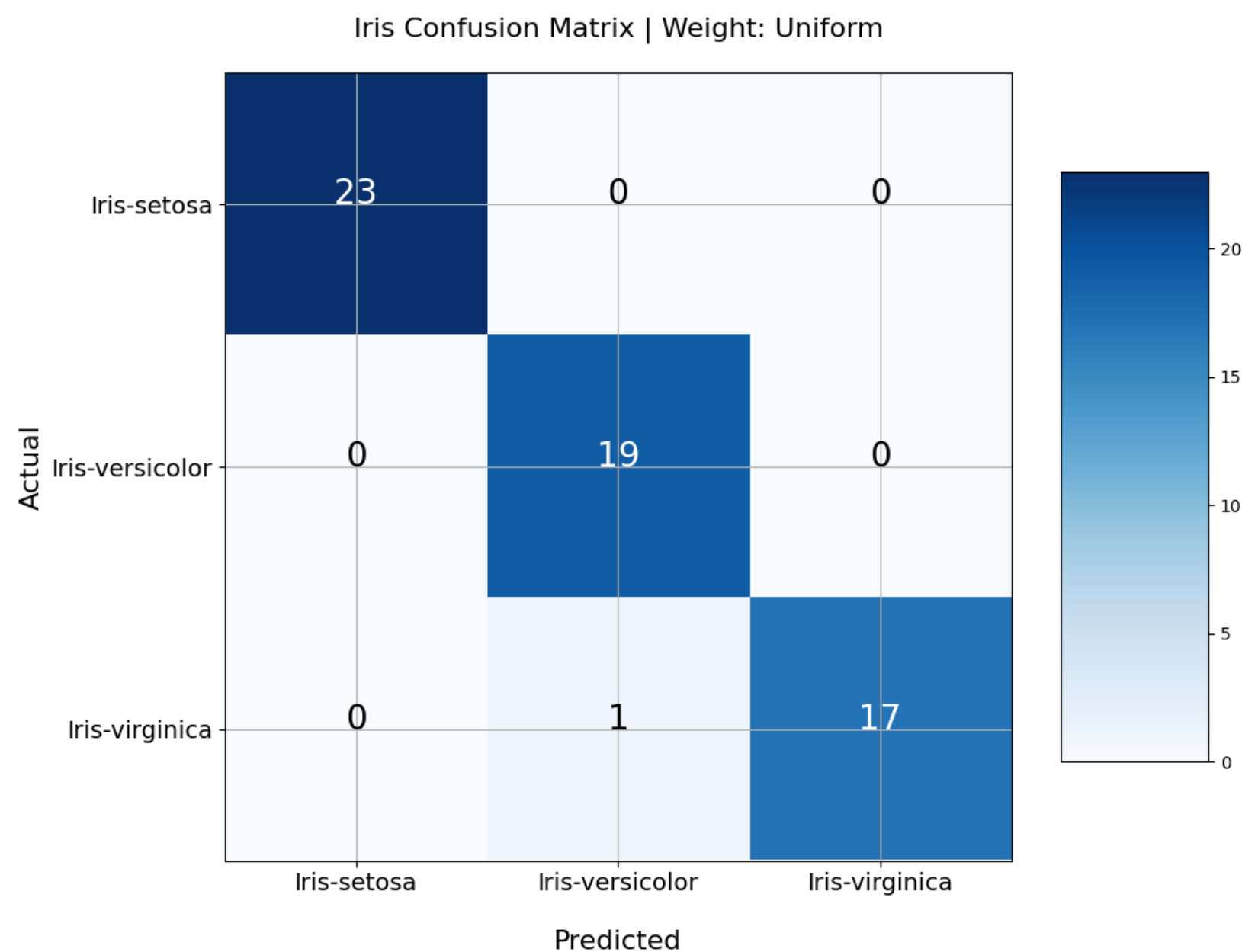
สัดส่วนข้อมูล 60:40

- Error Rate

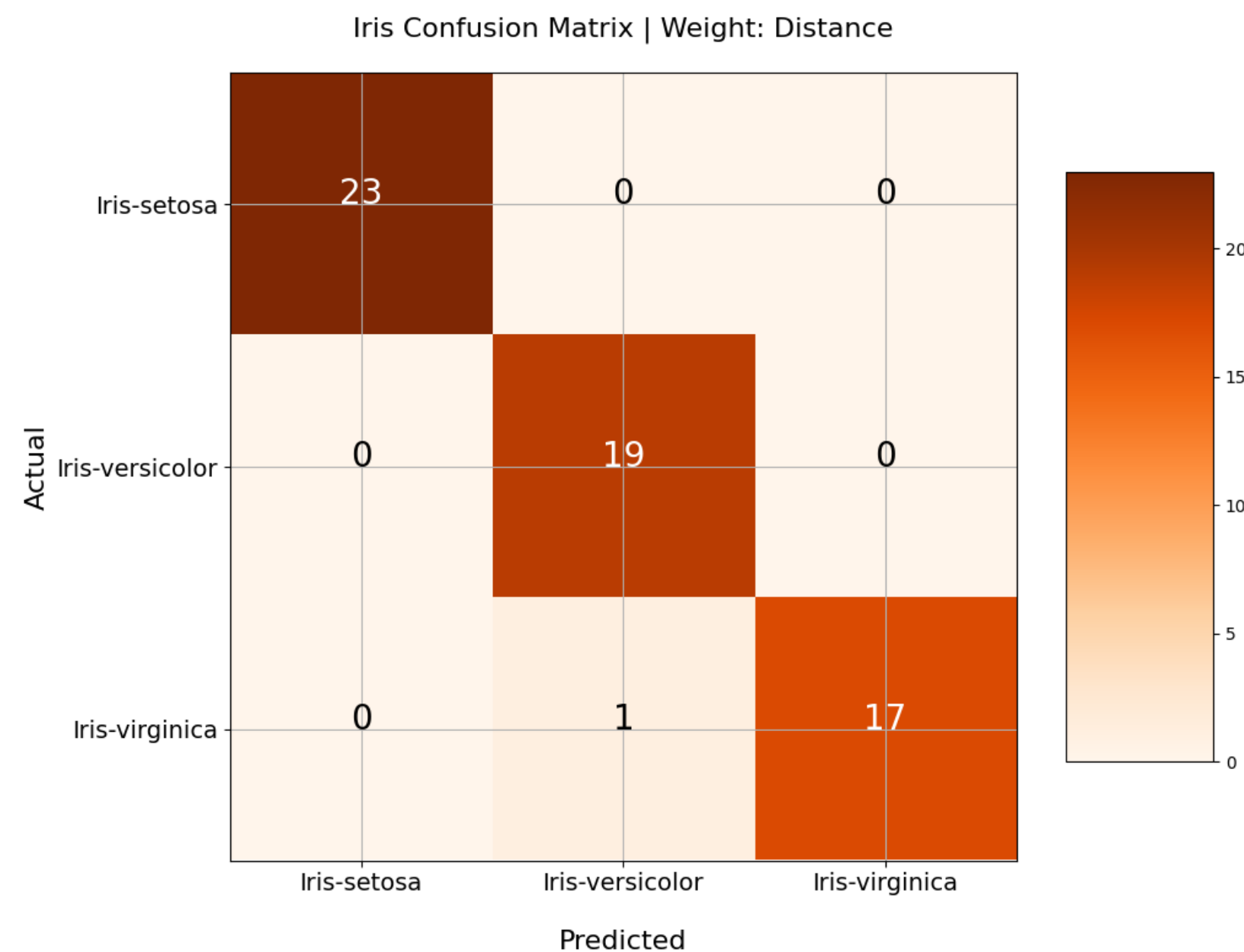


สัดส่วนข้อมูล 60:40

- ทดสอบด้วยค่า K ที่ Mean Error ต่ำสุด



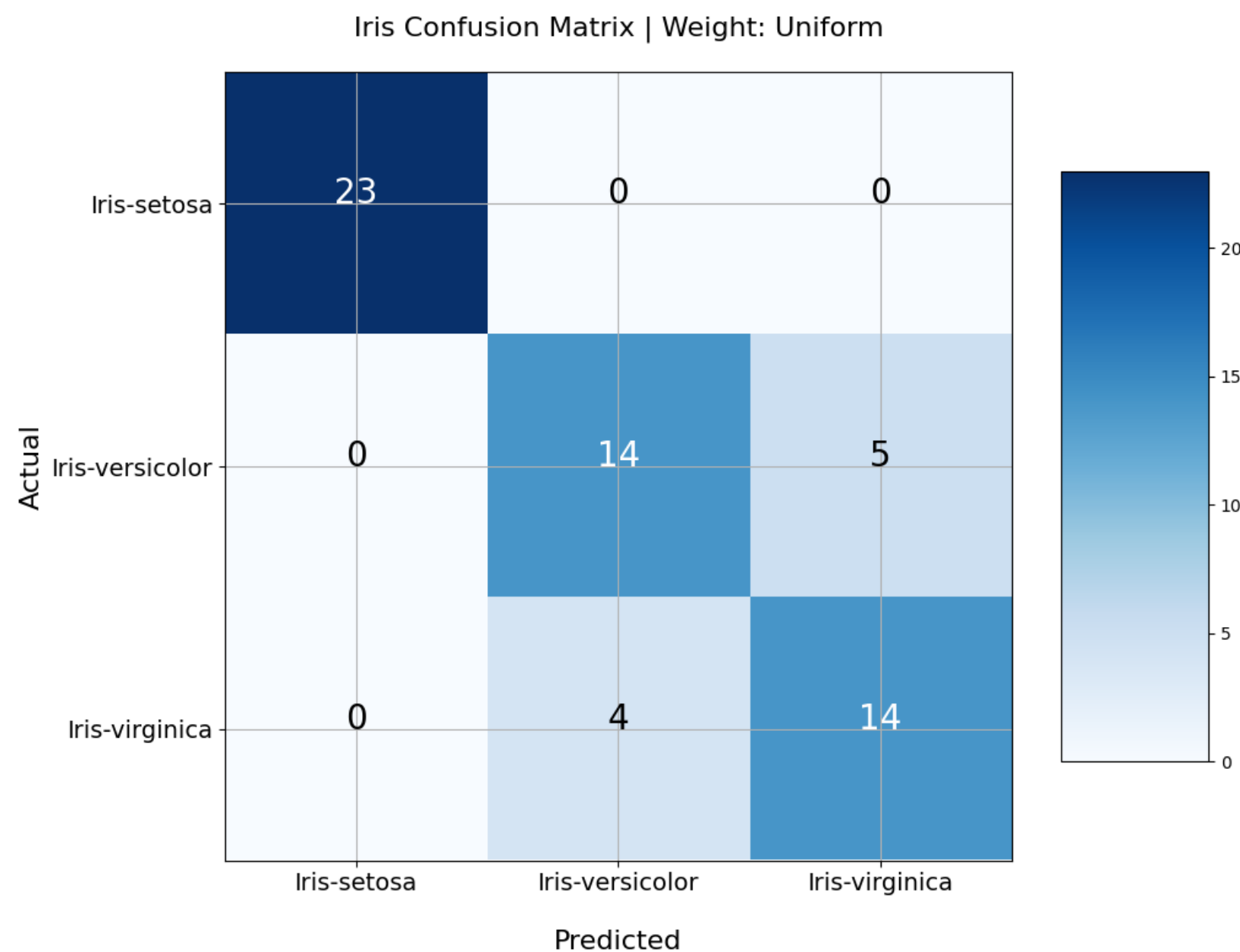
ความแม่นยำ: 98.33%



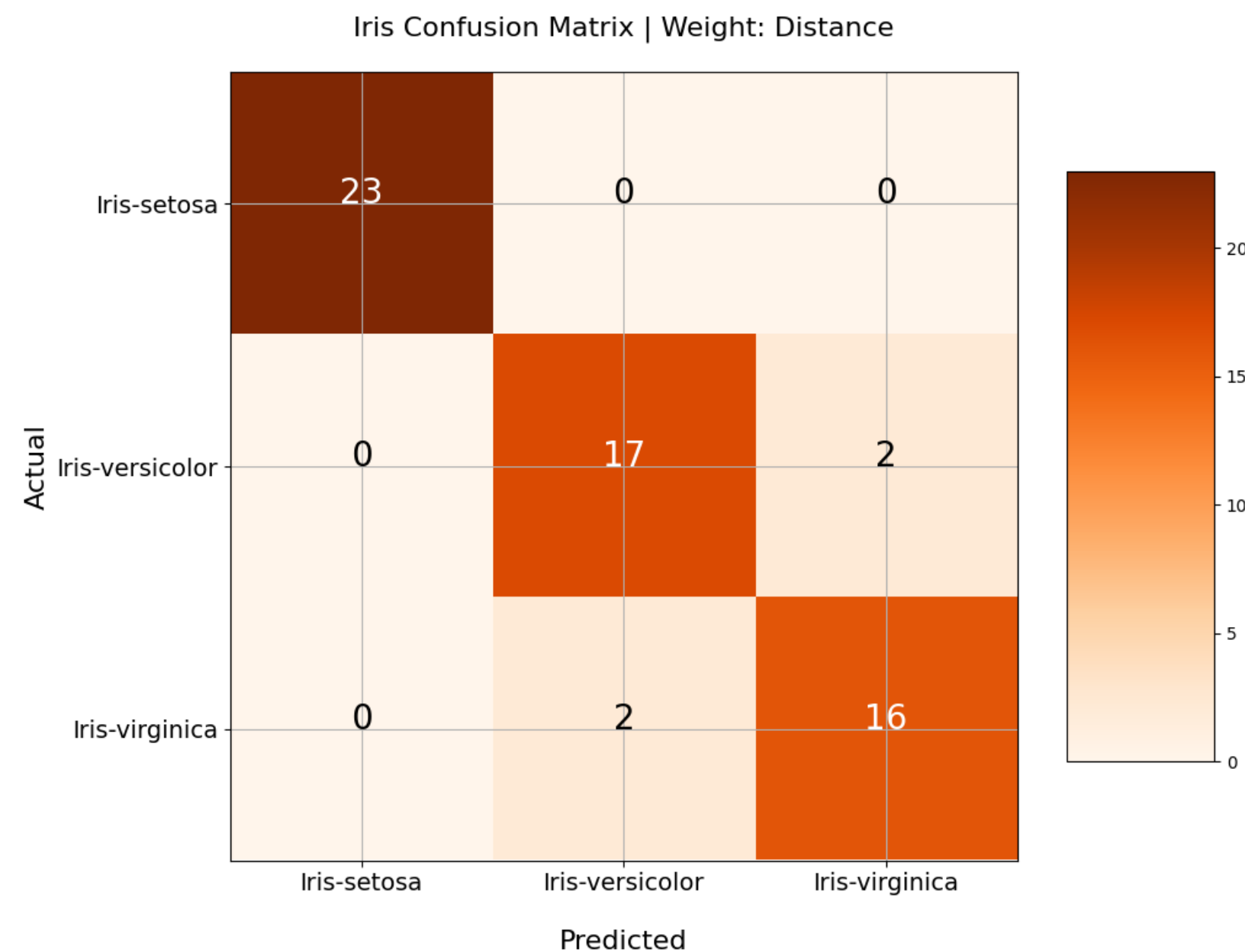
ความแม่นยำ: 98.33%

สัดส่วนข้อมูล 60:40

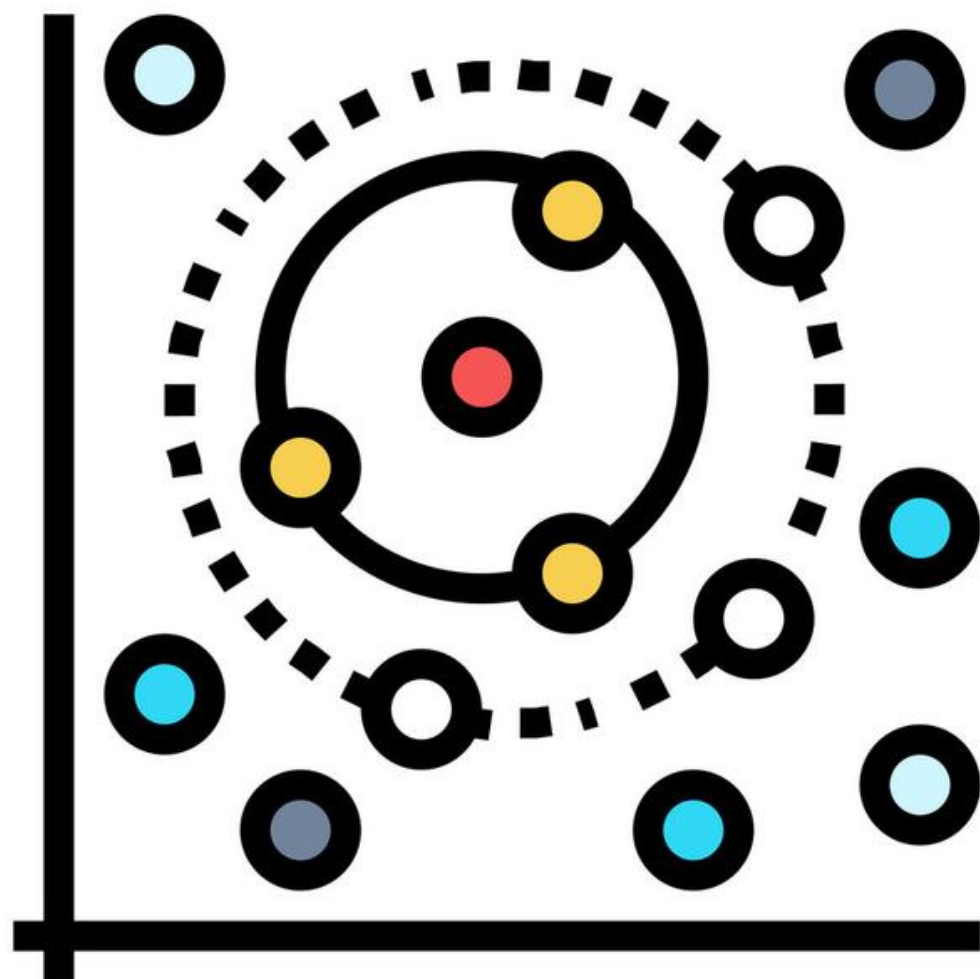
- ทดสอบด้วยค่า K ที่ Mean Error สูงสุด



ความแม่นยำ: 85.00%



ความแม่นยำ: 93.33%



ขอบคุณครับ

นายศุภณัฐ แซ่เตีย

ID: 6505000270

อ้างอิง

- scikit-learn developer. (2568). **Glossary of Common Terms and API Elements**. สืบค้นเมื่อ 23 กุมภาพันธ์ 2568, จาก https://scikit-learn.org/stable/glossary.html#term-random_state
- scikit-learn developer. (2568). **KNeighborsClassifier**. สืบค้นเมื่อ 23 กุมภาพันธ์ 2568, จาก <https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html>
- scikit-learn developers. (2568). **train_test_split**. สืบค้นเมื่อ 23 กุมภาพันธ์ 2568, จาก https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html
- vectorwin. (2568). **K nearest neighbors knn algorithm color icon vector image**. [Image]. สืบค้นเมื่อ 23 กุมภาพันธ์ 2568, จาก <https://www.vectorstock.com/royalty-free-vector/k-nearest-neighbors-knn-algorithm-color-icon-vector-50297687>