Proof. Since E is a cyclic $K[X]$-module, there is some $u \in E$ so that E is generated by $u, f(u), f^2(u), \ldots$, which means that every vector in E is of the form $p(f)(u)$, for some polynomial, $p(X)$. We claim that $u, f(u), \ldots, f^{n-2}(u), f^{n-1}(u)$ generate E, which implies that the dimension of E is at most n.

This is because if $p(X)$ is any polynomial of degree at least n, then we can divide $p(X)$ by $(X-\lambda)^n$, obtaining

$$p = (X-\lambda)^n q + r,$$

where $0 \leq \deg(r) < n$, and as $(X-\lambda)^n$ annihilates E, we get

$$p(f)(u) = r(f)(u),$$

which means that every vector of the form $p(f)(u)$ with $p(X)$ of degree $\geq n$ is actually a linear combination of $u, f(u), \ldots, f^{n-2}(u), f^{n-1}(u)$.

We claim that the vectors

$$u, (f-\lambda\mathrm{id})(u), \ldots, (f-\lambda\mathrm{id})^{n-2}(u)(f-\lambda\mathrm{id})^{n-1}(u)$$

are linearly independent. Indeed, if we had a nontrivial linear combination

$$a_0(f-\lambda\mathrm{id})^{n-1}(u) + a_1(f-\lambda\mathrm{id})^{n-2}(u) + \cdots + a_{n-2}(f-\lambda\mathrm{id})(u) + a_{n-1}u = 0,$$

then the polynomial

$$a_0(X-\lambda)^{n-1} + a_1(X-\lambda)^{n-2} + \cdots + a_{n-2}(X-\lambda) + a_{n-1}$$

of degree at most $n-1$ would annihilate E, contradicting the fact that $(X-\lambda)^n$ is the minimal polynomial of f (and thus, of smallest degree). Consequently, as the dimension of E is at most n,

$$((f-\lambda\mathrm{id})^{n-1}(u), (f-\lambda\mathrm{id})^{n-2}(u), \ldots, (f-\lambda\mathrm{id})(u), u),$$

is a basis of E and since $u, f(u), \ldots, f^{n-2}(u), f^{n-1}(u)$ span E,

$$(u, f(u), \ldots, f^{n-2}(u), f^{n-1}(u))$$

is also a basis of E.

Let us see how f acts on the basis

$$((f-\lambda\mathrm{id})^{n-1}(u), (f-\lambda\mathrm{id})^{n-2}(u), \ldots, (f-\lambda\mathrm{id})(u), u).$$

If we write $f = f - \lambda\mathrm{id} + \lambda\mathrm{id}$, as $(f-\lambda\mathrm{id})^n$ annihilates E, we get

$$f((f-\lambda\mathrm{id})^{n-1}(u)) = (f-\lambda\mathrm{id})^n(u) + \lambda(f-\lambda\mathrm{id})^{n-1}(u) = \lambda(f-\lambda\mathrm{id})^{n-1}(u)$$

and

$$f((f-\lambda\mathrm{id})^k(u)) = (f-\lambda\mathrm{id})^{k+1}(u) + \lambda(f-\lambda\mathrm{id})^k(u), \qquad 0 \leq k \leq n-2.$$

But this means precisely that the matrix of f in this basis is the Jordan block $J_n(\lambda)$. $\square$